



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

EĞİLİM PUANLARI KULLANILARAK ABİDE ÇALIŞMASINDAKİ MADDELERİN
DEĞİŞEN MADDE FONKSİYONU AÇISINDAN İNCELENMESİ

Mehmet YILMAZ

Yüksek Lisans Tezi

Ankara, 2021

Liderlik, arařtırma, inovasyon, kaliteli eđitim ve deđiřim ile

Daha ileriye ... En İyiyeye ...



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

EĞİLİM PUANLARI KULLANILARAK ABİDE ÇALIŞMASINDAKİ MADDELERİN
DEĞİŞEN MADDE FONKSİYONU AÇISINDAN İNCELENMESİ

INVESTIGATION OF DIFFERENTIAL ITEM FUNCTIONING OF THE TEST
ITEMS IN THE ABIDE STUDY BY USING PROPENSITY SCORES

Mehmet YILMAZ

Yüksek Lisans Tezi

Ankara, 2021

Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

Mehmet YILMAZ'ın hazırladıđı “Eđilim Puanları Kullanılarak Abide alıřmasındaki Maddelerin Deđiřen Madde Fonksiyonu Aısından İncelenmesi” bařlıklı bu alıřma j¼rimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitimde Ölme ve Deđerlendirme Bilim Dalında Y¼ksek Lisans** olarak kabul edilmiřtir.

J¼ri Bařkanı	Do. Dr. Melek G¼lřah řAHİN	İmza
J¼ri Üyesi (Danıřman)	Dr. Öğr. Üyesi K¼bra ATALAY KABASAKAL	İmza
J¼ri Üyesi	Dr. Öğr. Üyesi Sevda etin	İmza

Bu tez Hacettepe Üniversitesi Lisans¼st¼ Eđitim, Öğretim ve Sınav Yönetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri üyeleri tarafından / / tarihinde uygun gör¼lm¼ř ve Enstit¼ Yönetim Kurulunca / / tarihi itibarıyla kabul edilmiřtir.

Prof. Dr. Selahattin GELBAL
Eđitim Bilimleri Enstitüsü M¼d¼r¼

Öz

Bu çalışmanın amacı, ABİDE 2016 uygulamasının Türkçe, matematik ve fen bilimleri alt testlerinde yer alan maddelerin okul türü değişkenine göre değişen madde fonksiyonu (DMF) gösterip göstermediğini, eğilim puanı eşleştirme (EPE) öncesi ve sonrası incelemektir. Bu bağlamda EPE'nin DMF üzerindeki etkisi incelenirken ve EPE metodolojisinin adımları sunulmuştur. DMF analizleri, EPE öncesinde ve sonrasında Mantel-Haenszel (MH) ve lojistik regresyon (LR) yöntemleriyle gerçekleştirilmiştir. Çalışmanın verilerini Millî Eğitim Bakanlığı Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü tarafından sağlanan ABİDE 2016 uygulamasına katılan 6330 öğrenci oluşturmuştur. EPE analizleri öncesi tüm grup üzerinden gerçekleştirilen DMF analizleri sonucunda MH yöntemi ile dört, LR yöntemi ile bir maddenin DMF gösterdiği belirlenmiştir. EPE, en yakın komşuluk eşleştirme yöntemi ile gerçekleştirilmiştir. Eğilim puanı hesaplanmasında ortak değişkenler olarak sosyoekonomik düzey, aile ilgisi, anne ve baba eğitim düzeyi değişkenleri kullanılmıştır. Eşleştirme dengesi kontrol edilmiş ve eşleştirme sonucu ortak değişkenler bakımında devlet ve özel okul grubunda denge sağlandığı belirlenmiştir. EPE analizleri sonucu devlet ve özel okul grubundan 489 öğrenci eşleştirilmiş, DMF analizleri 978 öğrenci üzerinden gerçekleştirilmiştir. Bu analizler sonucunda MH yöntemi ile bir maddenin DMF gösterdiği, LR yöntemi ile DMF gösteren maddenin bulunmadığı belirlenmiştir. EPE sonrası DMF gösteren madde sayısında azalma olduğu görülmüştür. EPE kullanmanın, potansiyel DMF kaynakları üzerinde kontrol sağladığı sonucuna ulaşılmıştır.

Anahtar sözcükler: abide, değişen madde fonksiyonu, madde yanlılığı, eğilim puanı eşleştirme, ortak değişken.

Abstract

The purpose of this study is to examine whether the items in the Turkish, mathematics and science subtests of the ABIDE 2016 research show differential item functioning (DIF) according to the school type variable before and after the propensity score matching (PSM). In this context, while examining the effect of PSM on DIF and the steps of the PSM methodology are presented. DIF analyzes were performed before and after PSM using Mantel-Haenszel (MH) and logistic regression (LR) methods. The data of the study consisted of 6330 students who participated in the ABIDE 2016 application provided by the Ministry of National Education, General Directorate of Measurement, Evaluation and Examination Services. As a result of DMF analyzes performed on the whole group before PSM analyzes, it was determined that four items with the MH method and one item with the LR method showed DMF. PSM was carried out by the nearest neighborhood matching method. The variables of socioeconomic level, interest in family, education level of mother and father were used as covariate in calculating the propensity score. The matching balance was checked, and as a result of the matching it was determined that the balance was achieved in the public and private school groups in terms of covariates. As a result of the PSM analyzes, 489 students from the public and private school groups were matched, and DIF analyzes were carried out on 978 students. As a result of these analyzes, it was determined that an item showed DMF by the MH method and that there was no item showing DIF with the LR method. It was observed that there was a decrease in the number of substances showing DIF after PSM. It has been concluded that using PSM provides control over potential sources of DIF.

Keywords: abide differential item functioning, item bias, propensity score matching, covariate.

Teşekkür

Tez çalışması sürecinde zorlu şartlara rağmen benimle kurduğu güçlü iletişim sayesinde tezimi bitirebilmemi sağlayan, çözüm odaklı ve olumlu tutumuyla elinden gelen tüm yardımı benden esirgemeyen, günümüzde daha çok ihtiyaç duyduğumuz sabır, nezaket ve anlayış gibi değerlere sahip olan ve önemseyen değerli hocam Dr. Öğr. Üyesi Kübra ATALAY KABASAKAL'a,

Tez savunma jürisindeki önerileri ile tezime yaptıkları önemli katkılar için değerli hocalarım Doç. Dr. Melek Gülşah ŞAHİN ve Dr. Öğr. Üyesi Sevda ÇETİN'e,

Ders dönemi boyunca edindiğim eşsiz deneyim ve bilgiler için değerli hocalarım Prof. Dr. Selahattin GELBAL, Prof. Dr. Hülya KELECİOĞLU ve Doç. Dr. Burcu ATAR'a,

Tez çalışmamı gerçekleştirebilmem için gerekli verileri sağlayan MEB Ölçme Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü'ne,

Hayatımın her aşamasında en büyük destekçilerim ve güç kaynağım, kıymetlerini her geçen gün daha iyi anladığım, huzur kaynağım anne ve babama,

Her başlangıcımı destekleyip bana her zaman güç veren, kendime olan inancımı her zaman canlı tutmamı sağlayan biricik eşim Merve YILMAZ'a çok teşekkür ederim.

Evlatlarım Bilge Esmâ ve Göktürk'e...

İçindekiler

Öz.....	ii
Abstract.....	iii
Teşekkür.....	iv
Tablolar Dizini.....	vii
Şekiller Dizini.....	ix
Simgeler ve Kısaltmalar Dizini.....	x
Bölüm 1 Giriş.....	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi.....	7
Araştırma Problemi.....	9
Sayıltılar.....	10
Sınırlılıklar.....	10
Tanımlar.....	10
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	11
Akademik Becerilerin İzlenmesi ve Değerlendirilmesi.....	11
Geçerlik.....	12
Yanlılık.....	13
Değişen Madde Fonksiyonu.....	14
Değişen Madde Fonksiyonu Belirleme Yöntemleri.....	17
Eğilim Puanı Eşleştirme.....	20
İlgili Araştırmalar.....	23
Bölüm 3 Yöntem.....	27
Araştırmanın Türü.....	27
Araştırma Evreni ve Örnekleme.....	27
Veri Toplama Süreci.....	27
Veri Toplama Araçları.....	28

Verilerin Analizi	29
Test ve Madde İstatistikleri.....	29
Faktör Analizi	32
Eğilim Puanı Eşleştirme	36
Bölüm 4 Bulgular ve Yorumlar.....	42
Alt Problem 1a'ya İlişkin Bulgular ve Yorumlar.....	42
Alt Problem 1b'ye İlişkin Bulgular ve Yorumlar.....	44
Alt Problem 1c'ye İlişkin Bulgular ve Yorumlar	45
Alt Problem 2a'ya İlişkin Bulgular ve Yorumlar.....	47
Alt Problem 2b'ye İlişkin Bulgular ve Yorumlar.....	49
Alt Problem 2c'ye İlişkin Bulgular ve Yorumlar	50
Alt Problem 3'e İlişkin Bulgular ve Yorumlar	52
Bölüm 5 Sonuç, Tartışma ve Öneriler	54
Sonuç.....	54
Tartışma.....	55
Öneriler	56
Kaynaklar	58
EK-A: MEB Veri Talebi Cevap Yazısı	69
EK-B: Etik Komisyonu Onay Bildirimi	67
EK-C: Etik Beyanı.....	69
EK-Ç: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu	70
EK-D: Thesis/Dissertation Originality Report.....	71
EK-E: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı	71

Tablolar Dizini

Tablo 1 ABİDE Çalışmasında Kullanılan Maddelerin Sayısı	7
Tablo 2 <i>i</i> Maddesi için 2x2'lik veri tablosu	17
Tablo 3 $\Delta\alpha_{MH}$ Değerlerini DMF Açısından Yorumlama	18
Tablo 4 $\Delta R2$ Değerlerinin Yorumlanması Amacıyla Belirlenen Kesme Noktaları .	20
Tablo 5 Örneklemeye Ait Frekanslar	27
Tablo 6 Madde İstatistikleri	30
Tablo 7 Okul Türü Değişkenine Göre Alt Testlerin Madde Güçlük Düzeyleri	31
Tablo 8 Okul Türüne Göre Test İstatistikleri	31
Tablo 9 KMO Katsayıları ve Bartlett Testi Değerleri	32
Tablo 10 Paralel Analiz Tablosu	33
Tablo 11 Alt Testlere Ait Özdeğerler ve Açıklanan Varyans Yüzdeleri	34
Tablo 12 Eşleştirme Öncesi Ortak Değişkenlere İlişkin T Testi Sonuçları	36
Tablo 13 Ortak Değişkenlere İlişkin Lojistik Regresyon Analizi Sonuçları	37
Tablo 14 Eşleştirme Sonrası Gruplardaki Kişi Sayıları	38
Tablo 15 Eşleştirme Sonrası Ortak Değişkenlere İlişkin T Testi Sonuçları	40
Tablo 16 Eşleştirme Öncesi ve Sonrası SOF Değerleri, Eşleştirme Sonrası Varyans Oranları	40
Tablo 17 Türkçe Alt Testine Ait Okul Türü Değişkenine Göre MH Analizi Sonuçları	42
Tablo 18 Türkçe Alt Testine Ait Okul Türü Değişkenine Göre LR Analizi Sonuçları	43
Tablo 19 Matematik Alt Testine Ait Okul Türü Değişkenine Göre MH Analizi Sonuçları	44
Tablo 20 Matematik Alt Testine Ait Okul Türü Değişkenine Göre LR Analizi Sonuçları	45
Tablo 21 Fen Bilimleri Alt Testine Ait Okul Türü Değişkenine Göre MH Analizi Sonuçları	46
Tablo 22 Fen Bilimleri Alt Testine Ait Okul Türü Değişkenine Göre LR Analizi Sonuçları	47
Tablo 23 Türkçe Alt Testine Ait Okul Türü Değişkenine Göre MH Analizi Sonuçları	47

Tablo 24 <i>Türkçe Alt Testine Ait Okul Türü Değişkenine Göre LR Analizi Sonuçları</i>	48
Tablo 25 <i>Matematik Alt Testine Ait Okul Türü Değişkenine Göre MH Analizi Sonuçları</i>	49
Tablo 26 <i>Matematik Alt Testine Ait Okul Türü Değişkenine Göre LR Analizi Sonuçları</i>	50
Tablo 27 <i>Fen Bilimleri Alt Testine Ait Okul Türü Değişkenine Göre MH Analizi Sonuçları</i>	51
Tablo 28 <i>Fen Bilimleri Alt Testine Ait Okul Türü Değişkenine Göre LR Analizi Sonuçları</i>	51
Tablo 29 <i>EPE Öncesi ve Sonrası DMF Analizi Sonuçları</i>	52

Şekiller Dizini

Şekil 1. Yanlılık ve etki belirleme için kavramsal adımlar.	5
Şekil 2. TB DMF Gösteren Maddeye Ait MKE.....	16
Şekil 3. TBO DMF Gösteren Maddeye Ait MKE.	16
Şekil 4. Türkçe Alt Testine Ait Yamaç Birikinti Grafiği	35
Şekil 5 Matematik Alt Testine Ait Yamaç Birikinti Grafiği.....	35
Şekil 6. Fen Bilimleri Alt Testine Ait Yamaç Birikinti Grafiği.....	35
Şekil 7. Eğilim Puanı Eşleştirmeye İlişkin Jitter Grafiği.....	39
Şekil 8. Eşleştirme Öncesi ve Sonrası Gruplara Ait Eğilim Puanı Dağılımları.	41

Simgeler ve Kısaltmalar Dizini

MEB: Millî Eđitim Bakanlıđı

ABİDE: Akademik Becerilerin İzlenmesi ve Deđerlendirilmesi

DMF: Deđişen Madde Fonksiyonu

EPE: Eđilim Puanı Eşleştirme

MH: Mantel-Haenszel

LR: Lojistik Regresyon

TB DMF: Tek Biçimli Deđişen Madde Fonksiyonu

TBO DMF: Tek Biçimli Olmayan Deđişen Madde Fonksiyonu

SED: Sosyoekonomik Düzey

Bölüm 1

Giriş

Bu bölümde problem durumu, araştırma problemi, araştırmanın amacı ve önemi, sayıtlar, sınırlılıklar ve tanımlar yer almaktadır.

Problem Durumu

Günümüzde bireylerin bilişsel veya duyuşsal özelliklerinin belirlenebilmesi için testler geliştirilmekte ve bu testler sınav uygulamalarında, eğitim araştırmalarında sıklıkla kullanılmaktadır. Bu bağlamda, testlerden elde edilen sonuçlardan faydalanarak bireyler hakkında çeşitli kararlar verilmektedir. Bireylerin başarı veya yetenek düzeylerine göre sıralanıp ortaöğretim ve yükseköğretime geçişte puan üstünlüğüne göre yerleştirilmeleri, kamu kurumlarındaki kadrolar için seçim yapılması, öğrenme güçlüklerinin ve ders başarısının belirlenmesi bu kararlardan bazılarıdır. Ayrıca eğitim araştırmalarında kullanılan testlerden elde edilen sonuçlara göre ülkeler; eğitim harcamaları, öğretim programları ve kısa/uzun vadedeki planları hakkında önemli kararlar almaktadır. Testlerden elde edilen sonuçlara göre verilen kararlar bireysel, sosyal ve politik sonuçlar ortaya çıkarabilir (Clauser ve Mazor, 1998). Bu yüzden kullanılan testlerin, hangi yapıyı ölçmek amacıyla geliştirildiyse o yapıyı yansıtan doğru ve hassas ölçümler yapması beklenir. Fakat eğitimdeki ölçmelerin çoğunlukla dolaylı ölçmeler olması ölçme sonuçlarına hata karışma olasılığını artırır (Tekin, 2010).

Bir değişkenin, başka bir değişken yardımıyla ölçülmesi dolaylı ölçmedir (Turgut ve Baykul, 2015). Başarı, yetenek ve zekâ gibi bilişsel özelliklerin ölçülmesi sürecinde yapılan ölçme de, ölçülmek istenen yapının niceliksel göstergelere dönüştürülmesi bakımından dolaylı bir ölçmedir. Zekâ, yetenek ve başarı gibi özellikleri göstergeler yardımıyla ölçen testlerin sonuçlarına farklı değişkenlerin etkisinin yansımaması için bu testler geliştirilirken ciddi çalışmalar yapılır. Fakat yine de test puanlarının, ölçülmesi amaçlanan yapıyla ilişkisiz özelliklerden etkilenmesi kaçınılmazdır (Crocker ve Algina, 1986). Ölçülmek istenen yapıyla ilişkisiz özelliklerin test puanlarına karışması, bu puanları kullanarak verilecek kararların uygunluğunu ve test puanlarının anlamlılığını olumsuz etkilediğinden geçerlik için bir tehdittir. Ölçülen yapıyla ilişkisiz yapıların test üzerindeki etkisini yok etmek

imkânsız olduğundan, en azından herhangi bir gruba yönelik avantaj ya da dezavantaj sağlamadığından emin olunmalıdır.

Bir testin adil olması, testteki maddelerin ve testin uygulama koşullarının herhangi bir gruba avantaj veya dezavantaj sağlamaması ile mümkündür. Ölçülen yapıyla ilişkisi olmayan grup özelliklerinin test puanlarını etkilemesi testin yanlı olmasına sebep olacak ve test sonuçlarının geçerliğini tehdit edecektir (Clauser ve Mazor, 1998). Farklı testler üzerine yapılan birçok çalışmada cinsiyet, kültür, dil ve sosyoekonomik düzey gibi değişkenlerin etkisi ve bir grupta olup diğerinde olmayan özellikten dolayı maddelerin yanlı çalışıp çalışmadığı incelenmiştir (Atalay, 2010; Ayan, 2011; Gümüş, 2018; Özmen, 2013; Suna, 2012; Ulutaş, 2012; Yıldırım, 2006; Yıldırım, 2017). Bu çalışmalarla testin ölçmeyi amaçladığı yapı dışında özellikleri ölçüp ölçmediği belirlenebilmekte ve testin geçerliğine kanıtlar sunulabilmektedir. Madde yanlılığı test geçerliğini tehdit edebileceğinden dolayı bu konudaki çalışmalar test geliştirme sürecine olumlu katkı sağlamaktadır.

Bir ölçme aracı, ölçmeyi amaçladığı özellikle ilgisi olmayan cinsiyete, etnik kökene, sosyoekonomik düzeye göre bazı gruplara avantaj ya da dezavantaj sağlıyorsa, aracın amacına uygun olduğu veya amacı için yeterli olduğu iddia edilemez (Miller, Linn ve Gronlund, 2009). Zumbo (1999) madde yanlılığını, teste katılan gruplardan birinin üyelerinin maddeye doğru yanıt verme olasılığının, testin amacıyla ilişkili olmayan; test maddesi veya testin uygulama ortamı gibi sebeplerden dolayı diğer grup üyelerine göre düşük olması şeklinde tanımlamıştır. Bu tanıma göre benzer yetenek düzeyindeki bireylerin bir maddeye doğru yanıt verme olasılıklarının, ölçülmek istenen özellikle ilişkisi olmayan grup üyeliğinden etkilenmesi, o maddede potansiyel bir yanlılık olduğuna işaret edebilir. Maddelerin yanlı olması, bazı gruplara avantaj veya dezavantaj sağladığından test sonuçlarına sistematik hata karışmasına neden olacaktır. Çünkü hatanın yönü ve miktarı, teste katılan bireylerin bulunduğu gruba göre farklılık gösterecektir (Osterlind, 1983). Bu sebeplerle bir testte yanlı madde bulunması, test sonuçlarının geçerliğini olumsuz etkilemektedir.

Bir maddenin yanlı olup olmadığı hakkında karar vermenin ilk aşaması değişen madde fonksiyonu (DMF) belirlemek amacıyla yapılan istatistiksel analizlerdir. Fakat maddenin DMF göstermesi, tek başına maddenin yanlı olduğuna karar vermek için yeterli değildir. DMF gösteren maddelerin yanlı olup olmadığına

uzman kanısına başvurarak karar verilebilir (Camilli ve Shepard, 1994). Ölçülmek istenen yapıda aynı yetenek düzeyinde olan fakat farklı gruplarda yer alan bireylerin (örneğin kızlar ve erkekler) bir maddeye doğru yanıt verme olasılıkları ait oldukları gruba göre önemli ölçüde farklılaşıyorsa DMF ortaya çıkmaktadır. Bu nedenle, DMF test puanlarının karşılaştırılabilirliğini tehdit etmektedir. Ayrıca bir maddenin DMF göstermesi o maddenin herhangi bir gruba yanlı davrandığına işaret edebilir. Bu sebeplerle DMF, bilimsel araştırmalarda önemli bir odak olmuştur (Wu ve ark., 2017).

DMF çalışmalarıyla ilgili ilk resmi çalışmalar 70'lerin sonları ile 80'lerin başlarında gerçekleştirilmiştir (Camilli, 2006). Başlangıçta "madde yanlılığı" araştırması olarak adlandırılan DMF, test geliştiriciler ve kullanıcılar için ahlaki bir zorunluluk olarak test adaletiyle ilişkilendirilmişti (Wu ve ark., 2017). Bu yıllarda, yanlılık ve DMF kavramları birbiri yerine kullanılabilmekteydi ve hangi kavramın kullanılacağı yalnızca bir tercih meselesiydi (Millsap ve Everson,1993). Holland ve Wainer (1993), DMF konusunun eksik ve yanıltıcı terimlerle dolu olduğunu belirtmiş ve DMF teriminin yanlılıktan ayırt edilmesi gerektiğini ortaya koymuştur. Zumbo (1999) ise DMF'nin yanlılık için zorunlu ancak yeterli olmadığını ileri sürmüştür.

DMF analizlerinde, bireyler karşılaştırılan grup özelliği bakımından referans ve odak grup olmak üzere çoğunlukla iki gruba ayrılır ve iki gruptaki bireyler yetenek düzeyine göre eşleştirilirler. Yetenek düzeylerine göre eşleştirilen bu bireylerin test maddelerindeki performansları istatistiksel yöntemlerle karşılaştırılır. Bir maddenin DMF gösterdiğinin istatistiksel yöntemlerle belirlenmesinin ardından maddenin yanlı olup olmadığına karar vermek için yargısal bir süreç başlar. Bu süreçte uzmanlar, madde üzerinde odak ve referans gruba atfedilebilir özelliklerden dolayı yanlılık kaynaklarını araştırır ve DMF'nin neden belirli maddelerde ortaya çıktığına nedenler arar. Ancak bu her zaman kolay olmayabilir. Angoff (1993), test geliştiricilerin mükemmel uygunluktaki maddelerin çok büyük miktarlarda DMF göstermesini anlamlandırmada sorun yaşadıklarını bildirmiştir. DMF kaynaklarının belirlenemediği durumlarda, öğrencilerin arka plan özellikleri gibi diğer ilgili faktörlerin kontrol edilmesi yararlı olacaktır. Öğrenci arka plan değişkenleri genellikle grup farklılıklarının ve DMF kaynaklarının potansiyel açıklamaları olarak kabul edilir (Arikan, Vijver ve Yağmur, 2018). Örneğin; okul türüne (özel-devlet) göre gerçekleştirilen bir DMF çalışmasında karşılaştırılan gruplar, sosyoekonomik düzey

değişkenine göre çoğunlukla benzer yapıda olmayabilir. Gerçekleştirilen bu çalışmada sosyoekonomik düzey ve başarı ilişkisi düşünüldüğünde DMF'nin esas kaynağının okul türünden ziyade sosyoekonomik düzey değişkenine atfedilmesi gerekebilir. Bu sebeple böyle bir araştırmada sosyoekonomik düzey önemli bir arka plan değişkenidir. Sosyoekonomik statü, cinsiyet, okul türleri, test motivasyonu, kültürel farklılıklar, öğrenme fırsatı DMF kaynağı olabilecek bağlamsal değişkenlerdendir (Van de Vijver ve Tanzer, 2004; Zumbo ve Gelin, 2005).

Arka plan değişkenleri bir testte yer alan bir maddenin cevaplanmasındaki farklılığın grup özelliklerine atfedilmesine engel olabilmektedir. Örneğin; eşit yetenek düzeyindeki özel ve devlet okulu öğrencilerinin matematik testindeki bir maddeye doğru yanıt verme olasılıklarının değişip değişmediğinin araştırıldığı bir çalışma düşünelim. Karşılaştırma okul türüne göre yapılıyor olsa bile test motivasyonu, öğrenme fırsatı öğrencilerin matematik başarısını etkileyebilecek değişkenlerdir. Her iki grupta da önceden var olan bu özellikler öğrencilerin yanıtlarındaki farklılığın okul türlerine atfedilebilir olmasını engelleyebilir. Arka plan değişkenleri, araştırma kapsamında dikkate alınmayan ancak araştırma sonuçlarını etkileyen ve gruplarda farklılık gösteren özelliklerdir. Bu özellikler; bağımlı değişkenin, bağımsız değişkenlere makul biçimde atfedilebilir olmasını engelleyebilir.

Arka plan değişkenlerini karıştırıcı değişkenler olarak ele alan Wu ve arkadaşları (2017), yanlılık araştırmalarında grupların rastgele seçiminin yetersiz olabileceğini vurgulamış ve araştırmaların daha geçerli sonuçlar üretebilmesi için gruplar arasındaki istenmeyen ortak değişken (karıştırıcı) dağılımlarının dengelenmesi gerekliliğini ele almıştır. Hatta yanlılık çalışmalarında ilk adımı karıştırıcı değişkenlerin dağılımının gruplar arasında dengelenip bundan sonra DMF analizlerinin gerçekleştirilmesi gerektiğini belirtmişlerdir. Wu ve arkadaşları (2017), yanlılık ve etki araştırmaları için Şekil 1'deki adımları önermişlerdir.



Şekil 1. Wu ve ark. (2017), yanılılık ve etki belirleme için kavramsal adımlar

Şekil 1'e göre yanılılık araştırması için ilk adım olarak ortak (karıştırıcı) deęişkenlerin incelenen gruplar üzerinde daęılımı dengelenmelidir. Aksi takdirde yanılılık ve etki incelenemez. Ortak deęişkenler dengelendikten sonra yanılılık belirleme çalışması yapılmalıdır, yanılılık belirleme çalışması yapılmazsa grup karşılaştırmaları hatalı olacaktır. Yanılılık belirleme çalışması sonrası yanılı çalışan maddeler üzerinde grup etkisi araştırılmalıdır.

Wu ve arkadaşları (2017), Arıkan, Vijver ve Yaęmur (2018), DMF kaynaęı olabilecek arka plan deęişkenlerin kontrolü için farklı gruplardaki bireylerin eğilim puanları ile eşleştirilmesini önermişlerdir. Bu şekilde yapılan eşleştirmede karıştırıcı deęişkenlerin daęılımı, DMF analizi gerçekleştirilen gruplar arasında dengelenmiş olacaktır. Eğilim puanı, saęlık bilimleri alanında yapılan deneysel çalışmalarda uygulanan deney etkilerinin tahminindeki yanılılığı azaltmak için Rosenbaum ve Rubin (1983) tarafından bir yöntem olarak önerilmiştir. Çünkü rastgele oluşturulmayan gözlemsel çalışmalarda deney ve kontrol grupları arasındaki ortak (karıştırıcı) deęişkenler, araştırmacı tarafından etkisi incelenen deney hakkında yanılı kararlar alınmasına neden olabilir. Eğilim puanlarına dayalı eşleştirmede

gruplardaki birimlerin eğilim puanları bu gruplar için ortak arka plan değişkenleri yardımıyla lojistik regresyon kullanılarak elde edilebilir (Thomas ve Rubin, 1996).

T.C. Millî Eğitim Bakanlığı (MEB), izleme ve değerlendirme çalışmaları için uluslararası boyuttaki Uluslararası Öğrenci Değerlendirme Programı (The Programme for International Student Assessment – PISA), Uluslararası Matematik ve Fen Eğilimleri Araştırması (Trends in International Mathematics and Science Study – TIMSS) eğitim araştırmalarına katılım sağlayarak kendi öğrenci başarısını diğer ülkelerin verileriyle karşılaştırma, güçlü ve gelişime açık yönlerini belirleme imkânı yakalamaktadır. Ayrıca MEB, izleme ve değerlendirme çalışmalarını ulusal boyutta da devam ettirmektedir. Bunun için 2014 yılından itibaren ulusal bir izleme ve değerlendirme sistemi çalışmaları başlatmış ve 2016 yılında Akademik Becerilerin İzlenmesi ve Değerlendirilmesi (ABİDE) çalışmasının ilk uygulamasını gerçekleştirmiştir (MEB, 2017). ABİDE çalışmasında 8. sınıf öğrencilerin okulda edindiği bilgileri günlük hayatta kullanabilme becerileri ölçülmeye çalışılmıştır. Ayrıca öğrencilerin üst düzey düşünme yeterlikleri, akademik başarılarıyla ilişkili duyuşsal, aile ve okul özelliklerinin belirlenmesi hedeflenmiştir. ABİDE çalışması becerilerin ölçülmesine odaklanması yönüyle PISA, kazanımları temel alması yönüyle de TIMSS ile benzerlik göstermektedir. Ancak hem il düzeyinde hem de ülke geneli olarak izleme olanağı sağlaması yönüyle, yalnızca ülke geneli izleme sağlayan PISA ve TIMSS'ten farklılık göstermektedir (MEB, 2017). Bu çalışmanın periyodik olarak iki yılda bir uygulanacak olması ülke ve iller düzeyinde mevcut politika ve faaliyetlerin etkililiğinin değerlendirilmesi için önemli bir veri olacaktır.

ABİDE çalışmasında öğrencilerin bilişsel özelliklerini belirlemek amacıyla Türkçe, matematik, fen bilimleri ve sosyal bilgiler alanlarında çoktan seçmeli ve açık uçlu maddelerin kullanıldığı testler geliştirilmiştir. Buna ek olarak öğrenci başarısıyla ilişkili olduğu düşünülen diğer değişkenlerle ilgili olarak öğrenci, öğretmen ve okul anketleri geliştirilmiştir. ABİDE çalışması, yaklaşık 38000 öğrenciden oluşan bir örneklem üzerinden yürütülmüştür (MEB, 2017). Türkçe, matematik, fen bilimleri, sosyal bilgiler testlerinin her birinde 20 madde yer almış, bazı maddelerin ortak kullanıldığı 12 kitapçık türü (A1-A4, B1-B4, C1-C4) oluşturulmuştur. ABİDE çalışmasında kullanılan madde sayılarının ve türlerinin alanlara göre dağılımı Tablo 1'de sunulmuştur.

Tablo 1

ABİDE Çalışmasında Kullanılan Maddelerin Sayısı

Alanlar	Çoktan Seçmeli	Açık Uçlu
Türkçe	13	14
Matematik	13	14
Fen Bilimleri	13	14
Sosyal Bilgiler	14	13

Bu çalışmada ABİDE çalışmasının 2016 uygulamasındaki Türkçe, matematik ve fen bilimleri alt testlerinde yer alan maddelerin okul türü (devlet-özel) değişkenine göre DMF gösterip göstermediği incelenmiştir. Bir sonraki aşamada eğilim puanlarına göre eşleştirme yaparak DMF analizleri tekrarlanmıştır. Eğilim puanlarına göre eşleştirme öncesinde ve sonrasında DMF analizlerinden elde edilen sonuçlar karşılaştırılmıştır.

ABİDE çalışmasındaki başarı testleri ve öğrenci anketleri ile elde edilen veriler kullanılarak alt testlerde başarıyı etkileyebilecek değişkenler eşleştirme değişkeni olarak kullanılmıştır. MEB (2017); anne eğitim düzeyi, sosyoekonomik düzey, eğitim hedefi, evdeki kitap sayısı, derse ilgi ve öz yeterlik algısını başarıyı etkileyen değişkenler olarak belirlemiştir. Bu çalışmada eşleştirme değişkeni olarak başarıyı etkileyen ve gruplar arasında anlamlı farklılık olan değişkenler kullanılmıştır. Luellen, Shadish ve Clark (2005), gruplar arasında istatistiksel olarak anlamlı farklılığın bulunduğu değişkenlerin ortak değişkenler olarak kullanıldığını belirtmişlerdir. Baser (2006), grup üyeliğini ve çıktı değişkenini etkileyen tüm değişkenlerin dahil edilmesi gerektiğini aksi halde hatalı eğilim puanları hesaplanacağını belirtmiştir.

Araştırmanın Amacı ve Önemi

Bu araştırmanın amacı, eğilim puanlarına göre eşleştirmenin DMF üzerindeki etkisini incelemektir. Bu bağlamda, ABİDE 2016 yılı uygulamasında kullanılan Türkçe, matematik ve fen bilimleri alt testinde yer alan maddelerin okul türü değişkenine göre DMF gösterip göstermediğini, öğrencileri eğilim puanlarına göre eşleştirmeden önce ve sonra incelenmiştir.

Neredeyse yarım yüzyıldır geniş ölçekli sınavlar üzerinde gerçekleştirilen çok sayıda DMF ve yanlılık çalışması bulunmasına rağmen DMF'nin ortaya çıkma nedenleri tam olarak belirlenememekte ve madde yazımı konusunda genel kabul gören öneriler sunulmamaktadır. Arıkan, Vijver ve Yağmur (2017), bu durumu yanlılık çalışmaları bulgularının tekrarlanabilirliğinin zayıf olmasına ve yanlılık kaynakları üzerinde yeterli kontrole sahip olmamaya bağlamaktadır. Eğilim puanları eşleşmesiyle, başarıyla ilgili arka plan değişkenlerinin dağılımının benzer olduğu gruplar üretilebilir. Bu sayede olası yanlılık kaynaklarının kontrolü sağlanarak ortaya çıkan DMF, odak ve referans grubu ayıran temel değişkene atfedilebilir. Bu araştırmada, eğilim puanlarına göre gruplar eşleştirilerek başarıyla ilgili olabilecek arka plan değişkenleri bakımından benzer gruplar oluşturulmuştur. Bu yönüyle bu araştırma olası yanlılık kaynakları üzerinde kontrol sağlayarak DMF analizlerinin gerçekleştirilmesi bakımından önemlidir. Ayrıca gerçek bir veri seti üzerinde eğilim puanlarına göre eşleştirmenin DMF analizlerine etkisinin incelenmesi ve ABİDE uygulamasında kullanılan maddelerin okul türü değişkenine göre DMF gösterip göstermediğinin incelenmesi de bu çalışmanın diğer bir önemidir.

Alanyazında ulusal ve uluslararası geniş ölçekli sınavlar üzerinde gerçekleştirilmiş çok sayıda DMF ve yanlılık çalışması bulunmaktadır (Atalay, 2010; Akcan, 2018; Ayan, 2011; Erdem, 2015; Gümüő, 2018; Özmen, 2013; Suna, 2012; Ulutaő, 2012; Yıldırım, 2006; Yıldırım, 2017). Bu çalışmalar, çoğunlukla testlerde yer alan maddelerin DMF gösterip göstermediğini belirleme, analiz yöntemlerini karşılaştırma ve yanlılık çalışmalarıyla olası yanlılık kaynaklarını belirleme amacıyla gerçekleştirilmiştir. Bu çalışmada, grupları eşleştirmek için eğilim puanlarının kullanılmasının, olası yanlılık kaynakları üzerinde kontrol sağlayacağı söylenebilir. Ayrıca arka plan değişkenleri bakımından eşit gruplar oluşturan bir yöntem sunulmasının araştırmacılara katkı sağlayacağı düşünülmektedir.

Ulusal bir eğitim araştırması olan ABİDE iki yılda bir tekrarlanacak olup ilk uygulaması 2016 yılında, ikinci uygulama 2018 yılında gerçekleştirilmiş ve 2020 uygulaması pandemi sebebiyle ertelenmiştir. Bu çalışmanın sonuçlarının, eğitim politikalarını ve yatırımlarını etkileyebileceği düşünüldüğünde ABİDE'de kullanılan testlere yönelik geçerlik araştırmalarının önemi daha da artmaktadır. Alanyazında henüz ABİDE çalışmasına yönelik bir DMF ve yanlılık araştırmasına rastlanmamıştır. Bunun yanında alanyazında eğilim puanlarına dayalı az sayıda

DMF çalışması yer almaktadır (Arıkan, Vijver ve Yağmur, 2017; Lee ve Geisinger, 2014; Martinková, Hladká ve Potužníková, 2020; Wu ve ark., 2017). Bu sebeplerle bu çalışmanın test geliştiricilere ve araştırmacılara katkı sağlayacağı düşünülmektedir.

Bu çalışmada, DMF analizleri Mantel-Haenszel (MH) ve lojistik regresyon (LR) yöntemleriyle gerçekleştirilmiştir. MH ve LR yöntemlerinin 100 kişi kadar küçük örneklerde bile DMF belirlemede yeterli güçte olduğu görülmüştür (Rogers ve Swaminathan, 1993). Daha doğru sonuçlar elde edebilmek amacıyla birden fazla yöntem kullanarak DMF analizleri gerçekleştirilmiş çalışmalar literatürde yer almaktadır (Akın Arıkan, Uğurlu ve Atar, 2016; Atalay, 2010; Doğan ve Öğretmen, 2008; Erdem, 2015; Gök, Kelecioğlu ve Doğan, 2010; Yıldırım, 2017). Bu nedenlerle farklı yöntemler kullanarak DMF gösteren maddelerin belirlenmesi, araştırma sonuçlarının güvenilirliğini artırmaktadır.

Araştırma Problemi

ABİDE araştırmasının test uygulamasında kullanılan maddeler eğilim puanları eşleştirmesi öncesi ve sonrasında okul türü değişkenine göre DMF göstermekte midir?

Alt problemler. Üçüncü Araştırmada üç alt problem bulunmaktadır.

- 1) ABİDE alt testlerinde bulunan maddeler okul türüne göre MH ve LR yöntemleriyle yapılan analiz sonuçlarına göre DMF göstermekte midir?
 - a. ABİDE Türkçe alt testinde bulunan maddeler okul türüne göre DMF göstermekte midir?
 - b. ABİDE matematik alt testinde bulunan maddeler okul türüne göre DMF göstermekte midir?
 - c. ABİDE fen bilimleri alt testinde bulunan maddeler okul türüne göre DMF göstermekte midir?
- 2) ABİDE alt testlerinde bulunan maddeler okul türüne göre eğilim puanları eşleştirmesi sonrasında MH ve LR yöntemleriyle yapılan analiz sonuçlarına göre DMF göstermekte midir?

- a. ABİDE Türkçe alt testinde bulunan maddeler okul türüne göre DMF göstermekte midir?
 - b. ABİDE matematik alt testinde bulunan maddeler okul türüne göre DMF göstermekte midir?
 - c. ABİDE fen bilimleri alt testinde bulunan maddeler okul türüne göre DMF göstermekte midir?
- 3) Eğilim puanları eşleştirme öncesi ve sonrası gerçekleştirilen DMF analizi sonuçları arasında farklılık bulunmakta mıdır?

Sayıtlılar

Adayların testlere ve anketlere verdiği yanıtların gerçek durumlarını yansıttığı varsayılmıştır.

Sınırlılıklar

Bu araştırma MEB tarafından 2016 yılında gerçekleştirilen ABİDE çalışmasının Türkçe, matematik ve fen bilimleri alt testinde yer alan çoktan seçmeli maddeler ile sınırlandırılmıştır. Ayrıca eğilim puanlarının hesaplanmasında kullanılan ortak değişkenler, ABİDE 2016 çalışmasının anket uygulamalarından elde edilen değişkenlerle sınırlıdır.

Tanımlar

ABİDE: 8. sınıfta öğrenim gören öğrencilerin örgün eğitimde kazanmış oldukları bilgi ve becerileri kullanabilme yeterliklerinin ölçüldüğü bir durum belirleme çalışmasıdır (MEB, 2017).

Eğilim Puanı: Gözlenen ortak değişkenlere göre belirli bir deney grubuna atanmanın koşullu olasılığı olarak tanımlanmaktadır (Rosenbaum ve Rubin, 1983).

Değişen Madde Fonksiyonu: Madde ile ölçülmek istenilen özelliğin aynı düzeylerinde eşleştirilmiş farklı gruplardan bireylerin maddeye doğru yanıt verme olasılıklarının farklılaşmasıdır (Zumbo, 1999).

Madde yanlılığı: Bir maddenin ölçmek istediği yapıyla ilgisiz değişkenlerin etkisinden dolayı bir gruba avantaj veya dezavantaj sağlamasıdır (Zumbo, 1999).

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Akademik Becerilerin İzlenmesi ve Değerlendirilmesi (ABİDE)

ABİDE çalışması, 8. sınıfta öğrenim gören öğrencilerimizin örgün eğitimde kazanmış olduğu bilgi ve becerileri kullanma yeterliklerinin ölçüldüğü bir durum belirleme çalışmasıdır. ABİDE, 2016 uygulaması sadece 8.sınıflarla gerçekleştirilmiş ancak 2018 yılı uygulamasında 4.sınıflar da çalışma kapsamına alınmıştır. ABİDE çalışmasıyla öğrencilerin okulda edindiği bilgileri günlük yaşamda ne ölçüde kullanabildiklerini ve zihinsel becerilere sahip olma durumlarını belirlemek amaçlanmaktadır. Bunun sonucunda da ulusal düzeyde dönütler verecek bir izleme ve değerlendirme sisteminin kurulması planlanmaktadır. ABİDE'nin genel amacı, 8. sınıf öğrencilerinin bilişsel becerilerinin ortaya konulması ve öğrencilerin başarılarıyla ilişkili olabilecek öğrenci, öğretmen ve okul özelliklerinin belirlenmesidir. Bunun yanında, öğrencilerin başarılarını etkileyen duyuşsal, aile ve okul özelliklerini ortaya koymak, üst yeterlik düzeyine ulaşan öğrenci sayısını ortaya koymak gibi alt amaçları da bulunmaktadır (MEB, 2017).

ABİDE çalışmasında öğrencilerin bilişsel özelliklerinin belirlenebilmesi için Türkçe, matematik, fen bilimleri ve sosyal bilgiler alanlarında geliştirilen çoktan seçmeli ve açık uçlu maddelerden oluşan testler kullanılmıştır. ABİDE uygulamasında 12 ayrı kitapçık (A1:A4, B1:B4, C1:C4) kullanılmıştır. A formu 9 maddeyle B ve C formuna, B formu da 9 madde ile C formuna bağlanmıştır. Sonuç olarak esas maddeler A, B ve C formunun bütün kitapçıklarında yer almaktadır. A1'den A4'e kadar bütün A kitapçıklarındaki maddelerin ilk 18 sorusu ortaktır. Aynı durum B ve C kitapçıkları için de geçerlidir. Her bir kitapçıkta yer alan maddelerin yaklaşık olarak yarısı açık uçlu, diğer yarısı ise çoktan seçmelidir. Yapılandırılmış açık uçlu maddeler 0-1, 0-2, 0-3 vb. şeklinde puanlanmıştır (MEB, 2017).

ABİDE çalışması kapsamında öğrenci başarısını belirleyen bağlamsal faktörler hakkında veri toplanmasının sağlanması amacıyla başarı testlerinin yanı sıra öğrenci, öğretmen ve okul anketleri hazırlanmıştır. Bu anketler; uygulamaya katılan bütün öğrencilere, bu öğrencilerin öğretmenlerine ve okul yöneticilerine uygulanmıştır (MEB, 2017).

Geçerlik

Geçerlik; testlerden elde edilen puanların anlamlarının ve kullanımlarının, değerlendirilmenin yapılaş amacına uygunluğu ve bu amaç için yeterliliğidir (Miller, Linn ve Gronlund, 2009). Böylece bir test herhangi salt bir geçerliğe sahip değildir. Aksine bir testten elde edilen puanlar bazı kullanımlar veya yorumlar için geçerli iken diğer kullanımlar ve yorumlar için geçerli değildir (Thorndike ve Thorndike-Christ, 2010). Bu tanımlara göre 8. sınıf matematik başarısını ölçme amacıyla geliştirilmiş yüksek geçerliğe sahip bir testten alınan puanların, lisedeki başarıyı yordama amacı için de yüksek geçerliğe sahip olduğu savunulmaz. Buna ek olarak fen okuryazarlığını değerlendirme amacıyla geliştirilmiş bir testten alınan puanların matematik işlem becerisinden etkilenmiş olması yine o testin geçerliğini olumsuz etkileyeceği söylenebilir.

Crocker ve Algina (1986) geçerliği, kapsam geçerliği, ölçüt dayanaklı geçerlik ve yapı geçerliği olmak üzere üçe ayırmaktadır. Eğitimde ve Psikolojide Ölçme Standartları (American Educational Research Association, 2014) ise geçerliği tekil bir kavram olarak ele almış ve geçerliğe farklı kanıt sağlama süreçleri üzerinde durmuştur. Bu süreçleri kapsama, ölçüte ve yapıya dair kanıt sağlama olarak belirlemiştir. Messick'e (1989) göre yapıya dair geçerlik kanıtı sağlama çalışmaları tüm kanıt sağlama çalışmalarını kapsamaktadır. Bu süreçler sonucunda geçerliğin tümüyle geçerli ya da geçersiz gibi bir önerme olmadığı derecelendirmeye açık bir olgu olduğu belirtilmiştir (Thorndike ve Thorndike-Christ, 2010).

Ölçme aracının; içerdiği görevlerin ölçüldüğü iddia edilen alanı iyi temsil edememesi, ölçülen alanın en önemli kısımlarının örnekleyememesi, aynı değişkeni ölçen başka ölçme araçları ile arasındaki ilişkinin zayıf olması, adaylara sağladığı puanın ölçüldüğü iddia edilen özelliğin anlamlı bir göstergesi olmaması gibi sonuçların her biri geçerlik için birer tehdittir.

Yansızlık kavramı geçerliğe dahildir. Ölçme işlemi, kadınlara, azınlık gruplarına, anadili farklı olanlara, düşük gelir gruplarına haksızlık yapıyorsa amaca uygun olduğu veya amaç için yeterli olduğu iddia edilemez (Miller, Linn ve Gronlund, 2009). Test puanlarının adil ve geçerli bir şekilde yorumlanmasına yönelik en büyük tehdit, testi alan gruplar için puanları sistematik olarak düşüren ya da yükselten ve amacına uygun olmayan puanlama sonuçlarına neden olan ölçülen yapıyla alakasız

varyans üretebilecek boyutlardır (American Educational Research Association, 2014). Örneğin; bir matematik başarı testinin cinsiyete göre herhangi bir gruba avantaj veya dezavantaj sağlaması testten alınan puanların geçerliği için bir tehdittir. Öğrencilerin puanlarının gerçek matematik başarılarına göre farklılaşması beklenirken cinsiyetlerine göre de bu farklılığın ortaya çıkması sistematik hataya neden olacaktır.

Yanlılık

Alfred Binet, IQ testlerindeki bazı maddelerin zihinsel kapasiteden ziyade kültürel eğitim etkilerini ölçtüğünü düşünerek, düşük sosyoekonomik düzeyden çocuklarla çeşitli çalışmalar gerçekleştirmiş ve bu çalışmaların sonucunda bazı maddeleri testten çıkarmıştır. Bu çalışmaların, tarihsel süreçteki ilk yanlılık belirleme çalışmaları olduğu düşünülmektedir. Chicago Üniversitesinde 1950'li yıllarda Eells ve arkadaşları tarafından IQ testlerinin sonuçlarında, grupların test performansında gözlenen farklılığın sosyoekonomik düzey ve etnik geçmişe bağlı olup olmadığını anlamaya yönelik çalışmalar gerçekleştirilmiştir (Camilli ve Shepard, 1994). Bu çalışmaların ise modern yanlılık belirleme çalışmalarını başlattığı ve ileri taşıdığı öne sürülmüştür.

Eells ve arkadaşlarının çalışmaları ilerleyen süreçlerdeki çalışmalara dayanak noktası olmuş ve yanlılık belirlemeye yönelik istatistiksel analizler geliştirilmiştir. 1960 ve 1970'li yıllarda, yalnızca IQ testlerine değil eğitimsel testlere de yönelik yanlılık belirleme çalışmaları başlatılmıştır (Angoff, 1993). SAT (Scholastic Aptitude Test) sınavına katılan siyah ve beyaz adaylar için madde-grup etkileşimi incelenmesi, yanlılık belirleme sürecinde ilk formal çalışma kabul edilmektedir (Camilli, 2006).

Zumbo (1999) madde yanlılığını, teste katılan gruplardan birinin maddeye doğru yanıt verme olasılığının, testin ölçmek istediği yapıyla ilgisi olmayan özelliklerinden dolayı diğer gruplara göre daha düşük olması olarak tanımlamıştır. Yanlılık, yalnızca gruplar arasında puan farklılığının olması değildir. Eğer öyle olsaydı her madde bir alt gruba karşı yanlılık gösterebilirdi (Osterlind, 2002). Bu gerekçelerle bir maddenin yanlı olduğunu iddia edebilmek için, aynı yetenek düzeyinde eşleştirilmiş farklı gruptan bireylerin testin amacı dışında kalan

özelliklerinden dolayı maddeye doğru yanıt verme olasılıklarının farklı olması gerekmektedir.

Testten alınan puanların gruplara göre farklılık göstermesi gerçek yetenek düzeyindeki farklılıklardan kaynaklandığı durumlarda, maddenin yanlı olduğunu ileri sürmek doğru olmayacaktır. Bu yüzden madde etkisi ve yanlılık kavramlarının iyi anlaşılması gerekir. Madde etkisi, testin ölçmeyi amaçladığı özellik bakımından gruplar arasında gerçek yetenek düzeyi farklılığı olduğunda ortaya çıkar. Farklı gruplardan bireylerin bir maddeye doğru yanıt verme olasılığı farklı olabilir, bu farklılık grupların o maddeyle ölçülmek istenen gerçek yetenek farklılığından kaynaklanabilir (Zumbo, 1999). Bu durumda yanlılık kaynağı olarak bu farkı göstermek doğru olmayacaktır.

Camilli ve Shepard (1994) yanlılığı, test sonuçlarının sistematik hata içermesi olarak tanımlamıştır. Hatalı ölçüm sonuçları geçerliği olumsuz etkilediğinden yanlılık kavramı geçerlik araştırmalarının önemli bir konusu olmuştur. Yanlılık araştırmalarında karşılaştırılan alt evrenlerin, maddeden aldıkları puanların ortalamalarının farklılık göstermesini yanlılığa bağlamak uygun değildir. Bir madde, bir grup için nispeten daha zorsa ve bu zorluğun kaynağı testin ölçtüğü yapıyla ilgili değilse yanlı olduğu iddia edilebilir (Camilli, 2006). Bir maddenin yanlı olduğunu belirleyebilmek için ilk olarak DMF gösterip göstermediği istatistiksel olarak analiz edilmelidir (Zumbo, 1999). DMF, yanlılık kararı için doğrudan bir yol değil madde yanlılığının sadece bir indeksidir. DMF istatistikleri farklı gruplar için farklı işlev gösteren maddeleri tanımlamak için kullanılır; daha sonra tanımlanan maddelerin neden daha zor olduğuna dair mantıksal bir analizin yürütüldüğü yargısal süreç ile maddelerin yanlı olduğu belirlenir (Camilli, 2006).

Değişen Madde Fonksiyonu (DMF)

DMF, madde ile ölçülmek istenen yapının her bir yetenek düzeyinde eşleştirilmiş farklı grup üyelerinin maddeye doğru yanıt verme olasılığının farklılık gösterip göstermediğini belirleyen fonksiyondur (Zumbo, 1999). DMF belirleme yöntemleri yanlılık içermesi olası maddeleri belirlemek için kullanılır. Bir madde iki durumda DMF gösterebilir. Bunlardan ilki gruplar arasında, maddenin ölçtüğü

yapının gerçek yetenek düzeyinde farklılıktan ortaya çıkan madde etkisi, diğeri ise madde yanlılığıdır (Camilli ve Shepard, 1994).

DMF, madde yanlılığının bir indeksidir. DMF, yanlılık için gereklidir fakat yeterli değildir. DMF maddelerin farklı gruplarda nasıl işlediğini görmek için önemli bir yol göstericidir. Maddenin farklı gruplarda ölçülmek istenen yapıyla ilgisiz grup özelliklerinden dolayı farklı işleyiş göstermesi, isabetli bir değerlendirme için önemli bir sorundur. Bunun anlaşılabilmesi için DMF analizlerinin gerçekleştirilmesi gerekmektedir. DMF gösteren maddeler, maddenin cevaplanması için gerekli kazanımla ilgili olmayan özellikler barındırdığı için ölçülmek istenilen yapı ile ölçülen yapı arasındaki farkı ortaya koyma sürecinde DMF analizlerinin önemli sonuçlar ortaya koyduğu ifade edilmektedir. DMF, istatistiksel olarak belirlendikten sonra yargısal ve yoruma açık madde yanlılığı belirleme süreci başlar.

DMF analizlerinde, odak ve referans grubu olarak alınan iki gruptaki bireyler maddenin ölçtüğü yapının eşit yetenek düzeylerine göre eşleştirilir. Bu eşleştirilmenin yapılması için iç veya dış bir ölçüt kullanılır. Genellikle bu eşleştirme iç bir ölçüt olan toplam test puanına göre yapılır. Linn (1993), eşleştirme için kusursuz bir ölçüt olmadığını fakat test puanının iyi bir ölçüt olduğunu belirtmiştir (akt. Erdem, 2015). Eşleştirilen bireylerin test maddesindeki performansları istatistiksel yöntemler ile karşılaştırılır. Bu karşılaştırma sırasında benzer yetenek düzeyindeki farklı gruplara ait bireylerin, maddeye doğru yanıt verme olasılıklarının farklı olması maddenin DMF göstermesine neden olacaktır. DMF gösteren maddelerin yanlı olup olmadığı belirlemek için uzman kanısına başvurulur. DMF gösterdiği belirlenen bir maddenin, uzman bir ekip tarafından incelenmesi gereklidir. Bu uzman ekip performans farklılığına yol açan nedenin, ölçülen yapıyla ilişkili olup olmadığı kararını verecektir. Uzman ekibin gerçekleştirdiği bu yargısal süreç, yanlılığa ilişkin mantıksal kanıt toplanması olarak adlandırılır (Ayala, 2009).

DMF, tek biçimli (TB) ve tek biçimli olmayan (TBO) şekilde iki farklı biçimde ortaya çıkabilmektedir. Yetenek düzeyleri ile grup arasında hiçbir etkileşim olmadığı durumlarda TB DMF, yetenek düzeyleri ile gruplar arasında bir etkileşim olduğu durumlarda ise TBO DMF ortaya çıkmaktadır (Swaminathan ve Rogers, 1990). Yani bir madde yetenek düzeylerinin her seviyesinde bir gruba avantaj sağlıyorsa TB DMF'den, yetenek düzeylerinin belli bölümlerinde avantaj sağladığı grup farklılaşıyorsa TBO DMF'den söz edilebilir. Şekil 1 ve Şekil 2'de iki farklı grubun

madde karakteristik eğrileri (MKE) ile bir maddeye doğru yanıt verme olasılıkları sunulmuştur.



Şekil 2. TB DMF Gösteren Maddeye Ait MKE



Şekil 3. TBO DMF Gösteren Maddeye Ait MKE

Şekil 1'de MKE incelendiğinde maddenin yetenek düzeylerine göre değişmeksizin yalnızca bir gruba avantaj sağladığı ve yetenek düzeyleri boyunca bunun tutarlı bir şekilde devam ettiği, Şekil 2'de ise yetenek düzeyine göre maddenin avantaj sağladığı grubun değiştiği ve düşük yetenek düzeyleri için farklı gruba, yüksek yetenek düzeyleri için farklı gruba avantaj sağladığı görülmektedir.

Değişen Madde Fonksiyonu Belirleme Yöntemleri

Alanyazında farklı DMF yöntemlerinin varlığının yanı sıra bu yöntemlerin farklı sınıflandırmaları da mevcuttur. Bu sınıflandırmaların birinde yöntemler Klasik Test Kuramına (KTK) ve Madde Tepki Kuramına (MTK) dayalı olarak iki bölümde incelenir. Lojistik regresyon, ki-kare, varyans analizi, Mantel-Haenszel ve SIBTEST KTK'ya dayalı yöntemlerdir. Lord'un ki-kare testi, madde karakteristik eğrisi, olabilirlik oranı ve Raju'nun alan ölçümleri ise MTK'ya dayalı yöntemlerdendir.

Mantel-Haenszel ve lojistik regresyon yöntemlerinin; küçük örneklerde bile DMF belirlemede yeterli güçte olduğu görülmüştür. Ayrıca bu iki yöntemde de; odak ve referans gruplarının yetenek dağılımında farklar olduğunda, madde ayırt ediciliği yüksek veya çok zor ya da çok kolay maddeler varsa 1. Tip hata oranının benzer düzeyde arttığı sonucuna ulaşılmıştır (Rogers ve Swaminathan, 1993). Bu sonuçların yanında; ilgili araştırmalarda sıklıkla kullanılmaları, yazılım paketlerinin kolay erişilebilirliği ve pratik kullanımları sebepleriyle bu çalışma kapsamında Mantel-Haenszel ve lojistik regresyon yöntemi kullanılmıştır.

Mantel-haenszel. Mantel ve Haenszel (1959), hastalıklarla ilgili çalışmaların istatistiksel problemlerini incelemek için veri analizinde kullanılmak üzere 2x2'lik tablo yöntemini sunmuştur. Holland ve Thayer (1988), Mantel-Haenszel (MH) yöntemini DMF belirlemek için uygulamışlardır. Bu yöntemde eşleştirme ölçütü olarak gözlenen toplam test puanı kullanılır. Ayrıca referans ve odak olmak üzere iki grup düzeyi vardır.

MH, yorumlamasının kolay olması, küçük örneklerde kullanılabilmesi sebebiyle sıklıkla kullanılan DMF belirleme yöntemlerindendir. Analiz için odak ve referans gruptaki bireyler gözlenen toplam test puanına göre eşleştirilip her bir madde ve yetenek düzeyi için 2x2'lik tablolar oluşturulur.

Tablo 2

i Maddesi için 2x2'lik veri tablosu

Grup	Doğru	Yanlış	Toplam
Odak	A_i	B_i	n_{0i}
Referans	C_i	D_i	n_{1i}
Toplam	m_{1i}	m_{0i}	T_i

Her yetenek düzeyi için tablo oluşturulduktan sonra bu tablolar yardımıyla sabit bir olasılık oranı hesaplanır. Bu oran aşağıdaki denklem (1) ile gösterilir.

$$\alpha_{MH} = \frac{\sum_i A_i D_i / T_i}{\sum_i B_i C_i / T_i} \quad (1)$$

DMF için tek başına α_{MH} değeri yeterli değildir. Daha kolay yorum yapılabilmesi için logaritmik dönüşüm ile farklı bir ölçüğe dönüştürülerek $\Delta\alpha_{MH}$ elde edilir (Holland ve Thayer, 1988). Bu dönüşüm denklem (2) ile gösterilmiştir.

$$\Delta\alpha_{MH} = -2.35 \ln(\alpha_{MH}) \quad (2)$$

$\Delta\alpha_{MH}$ değeri 0'dan küçükse maddenin referans grup lehine, 0'dan büyükse maddenin odak grup lehine çalıştığı söylenebilir (Holland ve Thayer, 1988). Bu değer 0'a eşit ise madde DMF göstermemiştir. Zieky (1993), $\Delta\alpha_{MH}$ değerlerinin yorumlanması için standart bir sınıflama ölçütü önermiştir. Bu sınıflama ölçütü Tablo 3'te sunulmuştur.

Tablo 3

$\Delta\alpha_{MH}$ Değerlerini DMF Açısından Yorumlama

Değer	Düzye	DMF Düzeyi
$ \Delta\alpha_{MH} < 1$	A	Yok veya ihmal edilebilir düzeyde
$1 \leq \Delta\alpha_{MH} < 1,5$	B	Orta düzeyde
$1,5 \leq \Delta\alpha_{MH} $	C	Yüksek düzeyde

DMF analizleri sonucunda, B ve C düzey DMF gösteren maddeler DMF'li olarak işaretlenirken A düzey DMF gösteren maddeler dikkate alınmamaktadır (Zieky, 1993).

Lojistik regresyon. DMF belirleme yöntemleri arasında en etkili yöntemlerden biri lojistik regresyon (LR) yöntemidir (Camilli ve Shepard, 1994). Bu yöntemin DMF belirlemek amacıyla kullanılabileceği Swaminathan ve Rogers (1990) tarafından gösterilmiştir. LR yöntemi bir maddenin doğru cevaplanma olasılığı üzerine bir kestirim yapmaktadır. Bu sayede bireylerin madde üzerindeki performansı kestirilerek farklı gruptan bireylerin maddeye doğru yanıt verme olasılıklarının karşılaştırılmasına imkân verir. LR tekniği maddeye verilen yanıtları (1 veya 0) bağımlı; toplam puan, grup üyeliği (odak veya referans), toplam puan ve

grup etkileşimi ise bağımsız değişkenler olarak ele alır. Bir maddenin doğru cevaplanma olasılığını (p) kestiren LR modeli aşağıdaki gibidir.

$$p = \frac{e^z}{1+e^z} \quad (3)$$

$$z = \ln\left(\frac{p}{1-p}\right) \quad (4)$$

Denklem (3) ve (4)'ten hareketle aşağıdaki üç model kurulur. Bu modeller Denklem (5),(6) ve (7)'de verilmiştir.

$$z = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1\theta \quad (5)$$

$$z = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1\theta + \beta_2G \quad (6)$$

$$z = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1\theta + \beta_2G + \beta_3(\theta G) \quad (7)$$

Yukarıda denklem (5), (6) ve (7)'deki LR denklemlerinde θ bireyin yetenek düzeyini, G ise grup aidiyetini göstermektedir (Swaminathan ve Rogers, 1990). Bu yöntemde bireyin yetenek düzeyi ve farklı gruplara ait bireyleri eşleştirme ölçütü, toplam test puanı ile belirlenmektedir. β_1 , bireyin madde performansı ile bireyin yetenek düzeyi (toplam puan) arasındaki ilişkiyi; β_2 , gruplara göre madde performansındaki farkı, diğer bir ifadeyle referans ve odak grupların maddedeki ortalama performansları arasındaki farkı; β_3 ise bireyin yetenek düzeyi ile grup etkileşimini ifade etmektedir.

Lojistik regresyon yönteminde kurulan modelde üç denklem iki aşamalı olarak test edilir. Denklemler arasındaki ki-kare değerlerinin istatistiksel olarak anlamlı olup olmadığı incelenir. Birinci aşamada; Denklem (2) denklem (1)'e karşı test edilir. Denklem (2)'de, Denklem (1)'e ek olarak grup değişkeni eklenmiştir. Grup değişkeninin maddeyi doğru yanıtlama olasılığına etkisi araştırılarak Denklem (1) ve Denklem (2) arasındaki farkın anlamlı olması ($\beta_2 \neq 0$) tek biçimli DMF'ye işaret eder (Zumbo, 1999). İkinci aşamada; Denklem (3), Denklem (2)'ye karşı test edilir. Denklem (3)'te, Denklem (2)'den farklı olarak toplam puan ve grup etkileşimi değişkeni modele eklenir. Toplam puan ve grup etkileşimi modele dahil edildiğinde etkileşimin manidar olması ($\beta_3 \neq 0$) tek biçimli olmayan DMF'nin varlığını belirlemektedir (Zumbo, 1999).

DMF'nin varlığını, modellerin ki-kare (χ^2) değerlerine göre bu değerlerin farkının manidarlığını inceleyerek birbirine karşı test edilmesi ile belirlerken aynı zamanda her bir model için R^2 değerleri elde edilir. Model 3 ve Model 1'den elde edilen R^2 değerleri arasındaki farkı gösteren ΔR^2 değeri DMF'nin düzeyini belirlemek için kullanılır. ΔR^2 için Zumbo ve Thomas (1996), Jodoin ve Gierl (2001), Bakan Kalaycıoğlu ve Kelecioğlu (2011) tarafından önerilen kesme noktaları Tablo 4'te sunulmuştur.

Tablo 4

ΔR^2 Değerlerinin Yorumlanması Amacıyla Belirlenen Kesme Noktaları

DMF Düzeyi	Zumbo ve Thomas (1996)	Jodoin ve Gierl (2001)	Bakan Kalaycıoğlu ve Kelecioğlu (2011)
A	$\Delta R^2 < 0,13$	$\Delta R^2 < 0,035$	$\Delta R^2 < 0,010$
B	$0,13 \leq \Delta R^2 < 0,26$	$0,035 \leq \Delta R^2 < 0,070$	$0,010 \leq \Delta R^2 < 0,020$
C	$\Delta R^2 \geq 0,26$	$\Delta R^2 \geq 0,070$	$\Delta R^2 \geq 0,020$

DMF miktarları için A, ihmal edilebilir düzeyde; B, orta düzeyde; C, yüksek düzeyde DMF'yi göstermektedir (Zumbo ve Thomas, 1996; Jodoin ve Gierl, 2001; Bakan Kalaycıoğlu ve Kelecioğlu, 2011). Bu çalışmada, daha hassas ve duyarlı bir sınıflama olarak Bakan Kalaycıoğlu ve Kelecioğlu (2011) tarafından önerilen kesme noktaları kullanılmıştır.

Eğilim Puanı Eşleştirme

Deneyel araştırmalarda, değişkenler arasında oluşturulan neden-sonuç ilişkisini test etmek amacıyla araştırmacı işlem gruplarına seçkisiz atama yapmak, dışsal değişkenleri kontrol altına almak durumundadır. Sonuçların nedensellik bağlamında daha güçlü yorumlanabilmesi için örneklemden bireylerin işlem gruplarına seçkisiz (yansız) atanması ve dışsal değişkenlerin kontrolü önemli görülmektedir. Çünkü seçkisiz atama ile sonuç üzerinde potansiyel bir etkiye sahip tüm değişkenlerin eşdeğer olduğu ve grupların dengelendiği varsayılır (Büyüköztürk ve ark., 2018). Ancak bazı araştırmalarda seçkisiz atama yasal, etik ve uygulanabilirlik sebeplerinden dolayı mümkün olmayabilir (Luellen, Shadish ve Clark, 2005). Bu araştırmalarda, gruplar arasındaki farkların sadece deneysel

işlemden kaynaklandığının iddia edilebilmesi için deneklerin dışsal değişkene göre eşleştirilmiş iki gruba ayrılması gerekmektedir. Dışsal değişkenlerin kontrol edilmesinin, araştırmanın iç ve dış geçerliğine olumlu katkı yapacağı söylenebilir (Büyüköztürk ve ark., 2018).

Eğilim puanı eşleştirme (EPE), ilk olarak sağlık bilimlerinde gerçekleştirilen deneysel çalışmalarda deney ve kontrol gruplarının seçiminde yanlılığı önlemek amacıyla yöntem olarak önerilmiştir. Rosenbaum ve Rubin (1983), rastgele oluşturulmayan grupların ortak değişkenler bakımından sistematik olarak ayrılabilceğini ve bu sebeple doğrudan karşılaştırmaların hataya neden olabileceğini belirtmişlerdir. Deney ve kontrol gruplarını karşılaştırırken daha anlamlı sonuçlar ve işlem etkileri üzerine yapılan tahminlerde yanlılığı azaltmak amacıyla EPE yöntem olarak önerilmiştir.

EPE, biyomedikal alanda yapılan çalışmalarda ortaya çıkmış ve eğitim araştırmalarında kullanılması yavaş ilerlemiştir. Eğitim araştırmalarında grupların istatistiksel olarak eşdeğer hale getirilmesi ve araştırma bulgularından yapılan nedensel çıkarımların geçerliğine katkı sunması bakımından uygulanabilir ve faydalı bir tekniktir (Fan ve Nowell, 2011). Eğitim bilimlerinde gerçekleştirilen çalışmalarda da herhangi bir değişken bakımından grup karşılaştırmalarının anlamlı hale getirilmesi oldukça önemlidir. Örneğin; kız ve erkek öğrencilerin cinsiyetlerine göre başarı farklılıklarının incelenmesinde başarıyla ilişkili olabilecek arka plan değişkenlerinin her iki grup için de dengeli olması beklenir. Aksi halde gruplar arasındaki başarı farklılığını cinsiyet değişkenine atfetmek anlamlı olmayacaktır. Çünkü arka plan değişkenleri, araştırma kapsamında dikkate alınmasa dahi araştırma sonuçlarını etkileyebilecek, gruplarda önceden var olan ve gruplarda farklılık gösteren özelliklerdir. Bu özellikler, bağımlı değişkenin araştırma kapsamındaki bağımsız değişkenlere anlamlı biçimde atfedilebilir olmasını engelleyebilir.

Eğilim puanı, bazı değişkenler bakımında kişinin durumunun tahmin edilmesinde kullanılan koşullu olasılıktır ve 0 ile 1 arasında değerler alır. Eğilim puanı ($e(x_i)$) hesaplanması denklem (8) ile gösterilir (Rosenbaum ve Rubin, 1983):

$$e(x_i) = \text{prob}(z_i = 1|x_i) \quad (0 < e(x_i) < 1) \quad (8)$$

Bu denklemde z_i deęişkeni, deney ($z=1$) ve kontrol ($z=0$) grubunu göstermektedir. x_i ise gözlenmiş ortak deęişkenlerin vektörüdür.

Ortak deęişkenler (x_1, \dots, x_n) ve deney grubu (z_1, \dots, z_n) olmak üzere eğilim puanı, aşağıdaki denklem (9) ile hesaplanır.

$$prob(z_1, \dots, z_n | x_1, \dots, x_n) = \prod_{i=1}^N e(x_i)^{z_i} \{1 - e(x_i)\}^{1-z_i} \quad (9)$$

Eğilim puanı farklı yöntemlerle (probit model, diskriminant analizi, sınıflandırma ağaçları) kestirilebilmesine rağmen lojistik regresyon modeli en sık kullanılan ve pratik bir yöntemdir (Fan ve Nowell, 2011). Eğilim puanları için lojit dönüşümü kullanılırsa; aşağıdaki denklem (10) elde edilir.

$$\ln \frac{e(x_i)}{1-e(x_i)} = \ln \frac{prob(z_i = 1 | x_i)}{1-prob(z_i = 1 | x_i)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_N x_N \quad (10)$$

Lojistik regresyon modeli için eğilim puanları; denklem (11)'daki gibidir.

$$e(x_i) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_N x_N}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_N x_N}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_N x_N)}} \quad (11)$$

Olmos ve Govindasamy (2015), eğilim puanları hesaplandıktan sonra kontrol grubundaki katılımcılar arasından deney grubuna benzer eğilim puanı olanlar seçilerek devam edilir. Böylelikle deney ve kontrol grubundaki katılımcılar eşleştirilmiş olur. Rosenbaum (1989), eşleştirme yöntemlerini greedy ve optimal eşleştirme olarak iki ana hat üzerinde incelemiştir. Bu iki ana hat üzerinde farklı yöntemler yer almaktadır. Greedy eşleştirme yöntemleri eşleşen çiftler arasındaki farkın minimum olmasına odaklanırken, optimal eşleştirme yöntemleri tüm eşleşmedeki toplam farka odaklanır (Gu ve Rosenbaum, 1993). Alanyazında birçok eğilim puanı eşleştirme yöntemi bulunmaktadır (Arıkan, Vijver ve Yağmur, 2018). En yakın komşu eşleştirme, tam eşleştirme ve optimal eşleştirme yöntemleri eğitim alanındaki çalışmalarda kullanılmış yöntemlerdendir (Joldersma ve Bowen, 2010; Bowen, 2011; Lee ve Geisinger, 2014; Arıkan, Vijver ve Yağmur, 2018; Wu ve ark., 2017; Yılmaz Koğar, 2019; Chen, Liu ve Zumbo, 2019). En yakın komşuluk eşleştirme, greedy eşleştirme yöntemlerindenken; tam eşleştirme, optimal eşleştirme yöntemlerindedir. Greedy eşleştirmede, deney grubundaki birimler eğilim puanlarına göre sıralandıktan sonra her bir birim kontrol grubundaki en yakın eğilim puanına sahip birimle eşleştirilir (Austin, 2013). Optimal eşleştirmede ise temel yaklaşım eşleşen tüm birimler için toplam eğilim puanları farkının en küçük

olduğu birimleri bulmaktır (Rosenbaum, 1989). Başka bir deyişle eşleşen tüm birimler sonucunda ortalama mutlak eğilim puanı farkının minimum olması hedeflenir.

En yakın komşuluk eşleştirmesinin diğer yöntemlere göre daha az hesaplama gerektirdiği ve yanlılığı azaltmada oldukça başarılı olduğu belirlenmiştir (Rosenbaum ve Rubin, 1985; D'Agostino, 1998). Stone ve Tang (2013); lojistik regresyon modeli ile kestirilen eğilim puanlarının, en yakın komşuluk ve optimal eşleştirme yöntemleriyle eşleştirilmesinde benzer sonuçlar ürettiğini ve deney-kontrol grupları arasındaki standartlaştırılmış ortalama farkları azaltmada başarılı olduklarını belirlemişlerdir. Thoemmes ve Kim (2011), eğilim puanına yönelik çalışmalar en fazla kullanılan eşleştirme yönteminin en yakın komşu eşleştirme olduğunu belirlemişlerdir. Bu yöntemde deney grubundaki birimler eğilim puanlarına en yakın eğilim puanına sahip kontrol grubu birimleri ile eşleştirilir. Eşleştirme oranı 1:1 olarak belirlenirse her birim yalnızca bir birim ile eşleştirilir. Bu oran 1:N, N:1 ve N:N olarak belirlenirse farklı sayıda birimler birbiri ile eşleştirilebilir. Eşleştirme denklem (12)'deki şekilde gerçekleştirilir.

$$C(P_i) = \min_j ||P_i - P_j|| \quad (12)$$

P_i , deney grubundaki i. birimin; P_j , kontrol grubundaki j. birimin eğilim puanı göstermektedir. $C(P_i)$, deney grubunda i. birimin eğilim puanıyla kontrol grubunun j. biriminin eğilim puanı arasındaki farkların en küçüğüdür.

İlgili Araştırmalar

EPE analizlerine dayalı DMF incelenen araştırmalar. Martinková, Hladká ve Potužníková (2020), Çek Cumhuriyeti'nde akademik ve temel okula devam eden öğrencilerin öğrenme yeterliklerini karşılaştırmışlar ve bu karşılaştırmada öğrencilere uyguladıkları testlerin öğrencilerin devam ettiği okul türüne (akademik-temel) göre DMF gösterip göstermediğini incelemişlerdir. Akademik ve temel okula devam eden öğrencilere 6.sınıfta bir başarı testi uygulamışlar ve bu test sonuçlarına ek olarak aile eğitim düzeyi, sosyoekonomik düzey, okul dersleri başarısı, okulun bulunduğu şehrin büyüklüğü, akademik benlik algısı, cinsiyet değişkenlerine dayalı olarak EPE analizleri gerçekleştirmişlerdir. EPE analizleri en yakın komşu eşleştirme ile gerçekleştirilmiştir. Eşleştirilmiş grubun test verileri üzerinde yürütülen analizlerde iki maddenin DMF gösterdiğini belirlemişlerdir. Aynı öğrencilere 3 yıl

sonra 9.sınıfta aynı maddelere ek olarak daha zor maddelerin eklendiği bir test uygulanmış ve bu test sonucunda okul türüne göre öğrencilerin öğrenme yeterlikleri ilerlemelerinde anlamı bir farklılık görülmemiştir. Ayrıca 9.sınıfta uygulanan testte de iki maddenin DMF gösterdiği belirlenmiştir.

Arıkan, Van de Vijver ve Yağmur (2018), PISA 2012 matematik alt testinde eğilim puanları eşleştirme yöntemini kullanarak Endonezyalı, Türk, Avustralyalı ve Hollandalı öğrenciler arasında kültüre göre DMF gösterip göstermediğini ve madde etkisini incelemiştir. Eşleştirme değişkeni olarak sosyoekonomik düzey, öğrenme fırsatı ve cinsiyet değişkenlerini belirlemişlerdir. En yakın komşu, optimal ve tam eşleştirme yöntemlerini kullanarak EPE analizlerini gerçekleştirmişlerdir. Eğilim puanları eşleştirmesi sonrası gerçekleştirilen DMF analizlerinde orijinal veriden çok daha az DMF bulunduğu ve DMF kaynaklarını kontrol etmede eğilim puanları eşleştirme yönteminin uygun bir araç olduğu sonucuna ulaşılmıştır.

Wu ve ark. (2017), TIMSS 2007 matematik alt testinin İngilizce ve Fransızca versiyonunu alan öğrenciler arasında testin diline göre maddelerin DMF gösterip göstermediğini ve madde etkisini incelemiştir. Eşleştirme değişkeni olarak evdeki kitap sayısı, hesap makinesi kullanımı, ebeveynlerin eğitim düzeyi, bilgisayarın mevcudiyeti, matematik ödevine ayrılan zaman, matematiğe karşı tutum, matematiğe değer verme, özgüven ve okul güvenliği hakkındaki algı değişkenlerini kullanmışlardır. Optimal eşleştirme yöntemini kullanarak EPE analizlerini gerçekleştirmişlerdir. Eğilim puanlarına göre eşleştirme öncesi ve sonrasında gerçekleştirilen analizlerde DMF'li maddeler ortadan kalkmamış olsa da farklı sonuçlar elde edilmiş ve eğilim puanları eşleştirme yöntemi DMF analizlerinde ve madde etkisinin incelenmesinde ön koşul olarak sunulmuştur. DMF'nin gruplara atfedilebilmesi için EPE bir gereklilik olduğu belirtilmiştir.

Lee ve Geisinger (2014), Güney Kore'de uygulanan üniversite yeterlik sınavı İngilizce okuma testinde yer alan maddelerin cinsiyete göre DMF gösterip göstermediğini eğilim puanı eşleştirme yöntemi kullanarak incelemiştir. Eşleştirme değişkeni olarak İngilizce öğrenmeye başlama yaşı, yurtdışı dil eğitimi tecrübesi, İngilizce özel ders alma sayısı, İngilizce özel derse ayrılan zaman ve para, aile ilgisi, İngilizce öğrenmeye tutum değişkenlerini kullanmışlardır. Optimal eşleştirme yöntemini kullanarak EPE analizlerini gerçekleştirmişlerdir. MH ve LR yöntemlerinin her biri için eğilim puanı eşleştirmesi öncesi ve sonrasında gerçekleştirilen analiz

sonuçlarında DMF miktarları farklılık göstermiştir. DMF'li olarak işaretlenen madde sayısı EPE sonrası azalmıştır. EPE'nin potansiyel DMF kaynakları üzerinde kontrol sağladığını, geleneksel DMF anlayışına göre DMF'nin olası nedenlerini tahmin etmek için daha fazla bilgi sağladığı sonucuna ulaşmışlardır. EPE'ye dayalı DMF çalışmalarında DMF'nin sadece azalmadığını EPE sonrasında DMF miktarında artış ya da azalma olabileceği aktarılmıştır.

Joldersma ve Bowen (2010), dil sanatları okuryazarlığı testinde yer alan maddelerin testin İngilizce ve İspanyolca'sını alan öğrencilere göre DMF gösterip göstermediğini eğilim puanları eşleştirme yöntemi kullanarak incelemiştir. Eşleştirme değişkeni olarak sosyoekonomik düzey, toplam test puanı ve cinsiyet değişkenlerini belirlemişlerdir. Greedy eşleştirme yöntemini kullanarak EPE analizlerini gerçekleştirmişlerdir. Eşleştirme, ortak değişkenlerin aşamalı olarak algoritmaya eklendiği üç ayrı model ile gerçekleştirilmiştir. Modeller çoğunlukla tutarlı sonuçlar üretse de DMF düzeylerinde farklılıklar gözlenmiştir. DMF düzeylerinin farklılaşması modellerde kullanılan ortak değişkenlerin farklılaşmasından kaynaklandığı sonucuna ulaşılmıştır. Farklı dağılım özelliklerinin kontrolü, odak ve referans grup arasındaki DMF'nin anlaşılması ve tanımlanması konularında EPE'nin avantaj sağladığı belirtilmiştir.

Gizil sınıf analizlerine dayalı DMF incelenen araştırmalar. Yalçın (2017), TIMSS 2015 fen bilimleri alt testinde cinsiyete göre maddelerin DMF gösterip göstermediğini oluşturulan gizil sınıflarda incelemiştir. Analizler, iki aşamada gerçekleştirmiştir. İlk aşamada, öğrencilerin fene yönelik tutumu, fene karşı kendine güveni, fen derslerine katılımı ve fen öğrenmeye verilen değerine göre oluşan gizil sınıfları belirlemek için gizil sınıf analizi kullanmış ikinci aşamada, öğrencilerin duyuşsal özelliklerine göre oluşan gizil sınıflarda ve tüm grup için maddelerin DMF durumunu belirlemek için MH yöntemi kullanılmıştır. Tüm grup için yapılan DMF analizlerine göre ele alınan maddelerin sadece birinde DMF tespit edilmiştir. Ancak öğrenciler bazı duyuşsal özelliklerine göre gizil sınıflara ayrıldığında, bu maddenin iki gizil sınıfta DMF göstermediği, bir sınıfta ise farklı bir maddenin de DMF gösterdiği tespit etmişlerdir.

EPE'ye göre yürütülen araştırmalar. Rosenbaum ve Rubin (1983), deneysel çalışmalar için grupların atanmasında eğilim puanları eşleştirme yöntemiyle gerçekleştirilmesini önermişlerdir. Deneysel çalışmalarda grupların

sistematik olarak ayrılabilirliğini ve bu grupları karşılaştırmanın hatalı olabileceğini bu sebeple eşleştirilmiş grupları karşılaştırmanın daha anlamlı olacağını belirtmişlerdir.

Rosenbaum (1986), ABD'de lise öğrencilerinin okuldan ayrılma durumlarının bilişsel başarı testi puanları üzerindeki etkisini tahmin etmeye çalışmıştır. Bir okuldan örneklenen, okuldan ayrılan ve okula devam eden öğrencilerin çok değişkenli bir yöntem ile eşleştirilmesi sağlanmıştır. Okuldan ayrılma durumlarının başarı testi puanları üzerinde etkili olduğunu belirlemiştir.

Fan ve Novell (2011), eğitim araştırmalarında deneysel olmayan tasarım veya gözlemsel araştırma içeren araştırma durumlarında veya deneysel tasarımın yararlarının tam olarak gerçekleştirilmediği durumlarda nedensel çıkarımların geçerliliğini arttırmak için kullanılacak eğilim puan eşleştirme yöntemini tanıtmaktadır. Ayrıca çalışmalarında, eğilim puanları eşleştirme mantığını, temel uygulama adımlarını ve işlemleri bir veri örneği ile göstermişlerdir.

Yılmaz Koğar (2019), PISA 2015 Türkiye örnekleme verilerini kullanarak eğilim puanı eşleştirmesinin eğitim araştırmalarında kullanımını gösteren bir örnek sunmuştur. EPE analizleri öncesi ve sonrası özel okulda ve devlet okulunda öğrenim gören öğrencilerin başarılarını karşılaştırmıştır. Eşleştirme değişkeni olarak kültürel olanaklar, maddi varlık, evdeki eğitim kaynakları, evde sahip olunan eşyalar, bilgi iletişim teknolojileri kaynakları değişkenleri kullanılmıştır. EPE analizleri en yakın komşu eşleştirme yöntemini kullanarak gerçekleştirmiştir. Eğilim puanı eşleştirmesi ile grupların daha karşılaştırılabilir bir duruma getirdiği belirlenmiştir. Eğilim puanı eşleştirmesi öncesinde karşılaştırılan gruplar (özel ve devlet okulunda öğrenim görme) okuryazarlık puanları bakımından anlamlı farklılıklar gösterirken eşleştirme sonrasında okuryazarlık puanlarında gruplar arasında anlamlı farklılık bulunmamıştır.

Bu çalışmaların sonuçlarına göre EPE'nin, araştırmalarda nedensel çıkarımların geçerliliğine katkı sağladığı, bu sebeple eğitim bilimleri alanında yapılan çalışmalarda da bağımlı değişkenin bağımsız değişkene atfedilebilir olmasına katkı sağladığı söylenebilir. Buna bağlı olarak öğrenci başarısı üzerinde etkili olduğu düşünülen değişkenler üzerinde kontrol sağlayarak DMF kaynaklarını saptamada elverişli olduğu söylenebilir.

Bölüm 3

Yöntem

Bu bölümde araştırmanın türü, araştırmanın evren ve örnekleme, verilerin elde edilmesi, verilerin analizi ve bu analizlerde kullanılan bilgisayar programları üzerinde durulmuştur.

Araştırmanın Türü

Bu araştırmada, ABİDE 2016 yılı uygulamasında kullanılan Türkçe, matematik ve fen bilimleri alt testinde yer alan maddelerin okul türü değişkenine göre DMF gösterip göstermediğini, öğrencileri eğilim puanlarına göre eşleştirmeden önce ve sonra incelenmiştir. Var olan durumu ortaya çıkarması ve detaylı biçimde incelenmesi yönüyle nicel araştırmalar kapsamındaki betimsel model özelliği taşımaktadır.

Araştırmanın Evreni ve Örnekleme

Araştırmanın evrenini, ABİDE çalışmasının 2016 yılında gerçekleştirilen test uygulamasına katılan yaklaşık 38000 öğrenci oluşturmaktadır. Çalışmanın örneklemini, Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü'nün tarafından sağlanan verilere göre 6330 öğrenci oluşturmaktadır. DMF analizleri için değişkenlere göre öğrenci sayılarının dağılımı Tablo 5'te sunulmuştur.

Tablo 5

Örnekleme Ait Frekanslar

Değişken	Öğrenci Sayısı	
Okul Türü	Devlet	5791
	Özel	539

Okul türü değişkeni için 6330 kişilik örneklem üzerinden DMF analizleri gerçekleştirilmiştir.

Veri Toplama Süreci

Bu araştırmada kullanılacak veriler gerekli süreçlere uygun olarak Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü'nden tarafınca sağlanmıştır.

Bu veriler 6330 öğrencinin matematik testinde yer alan çoktan seçmeli maddelerden alınan puanları (1-0) ve anket verilerini kapsamaktadır. Çoktan seçmeli maddelerden alınan puanlar kullanılarak öğrencilerin Türkçe, matematik ve fen bilimleri testine ait toplam puan araştırmacı tarafından hesaplanmıştır.

Veri Toplama Araçları

Akademik Becerilerin İzlenmesi ve Değerlendirilmesi (ABİDE) çalışması, ortaokul 8. sınıfta öğrenim gören öğrencilerimizin örgün eğitimde kazanmış oldukları bilgi ve becerileri kullanma yeterliklerinin ölçüldüğü bir durum belirleme çalışmasıdır. ABİDE, 2016 uygulaması sadece 8.sınıflarla gerçekleştirilmiş ancak 2018 yılı uygulamasında 4.sınıflar da çalışma kapsamına alınmıştır.

ABİDE çalışmasıyla öğrencilerin okulda öğrendiklerini gündelik yaşamda ne ölçüde kullanabildiklerinin ve zihinsel becerilere sahip olma durumlarının belirlenmesi amaçlanmaktadır. Bunun sonucunda da ulusal düzeyde bir izleme ve değerlendirme sisteminin kurulması planlanmaktadır.

ABİDE araştırmasında genel amaç, 8. sınıf öğrencilerinin bilişsel becerilerinin ortaya koyulması ve öğrencilerin başarılarıyla ilişkili olabilecek öğrenci, öğretmen ve okul özelliklerinin belirlenmesidir. Bunun yanında, öğrencilerin başarılarını etkileyen duyuşsal, aile ve okul özelliklerini ortaya koymak, yeterlik düzeylerinin belirlenmesinden sonra düzenli olarak uygulanacak bu çalışma ile süreç içerisinde eğitimde gelişimi ve üst yeterlik düzeyine ulaşan öğrenci sayısını ortaya koymak gibi alt amaçları da bulunmaktadır (MEB, 2017).

ABİDE çalışması kapsamında öğrenci başarısını belirleyen ve etkileyen faktörlere yönelik veri toplanmasını sağlamak amacıyla başarı testlerinin (Türkçe, matematik, fen bilimleri, sosyal bilgiler alt testleri) yanı sıra öğrenci, öğretmen ve okul anketleri hazırlanmıştır. Öğrenci anketleri, uygulamaya katılan bütün öğrencilere; öğretmen anketleri, ABİDE uygulamasına katılan öğrencilerin dersine giren Türkçe, matematik, fen bilimleri ile sosyal bilgiler öğretmenlerine; okul anketleri ise ABİDE uygulamasının gerçekleştirildiği okulların müdürlerine uygulanmıştır (MEB, 2017).

Verilerin Analizi

Verilerin çözümlenmesi sürecinde, doğru yanıtlar için 1, yanlış yanıtlar ve boş bırakılan maddeler için 0 kodlanarak veri tamamen 1-0 şekline dönüştürülmüştür. Daha sonra alt testlerde yer alan maddelerden alınan puanlar toplanarak her bir öğrenci için Türkçe, matematik ve fen bilimleri test puanı hesaplanmıştır.

Alt testte yer alan maddeler için madde istatistikleri 6330 öğrenci üzerinden TAP programı (Brooks ve Johanson, 2003) ile hesaplanmış ve tablolaştırılmıştır.

DMF analizleri öncesinde, tek boyutluluğun sınanması amacıyla faktör analizi FACTOR 10.8.4 programı (Ferrando ve Lorenzo-Seva, 2013) ile gerçekleştirilmiş ve her bir alt testin tek boyutluluk varsayımlarını karşılayıp karşılayamadığı incelenmiştir.

EPE analizleri R yazılımında MatchIt paketi (Ho ve ark., 2007) ile gerçekleştirilmiş ve yine R yazılımında cobalt (Grefier, 2020), tableone (Yoshida, 2020) paketleri ile eşleştirme dengesi kontrol edilmiştir.

Maddelerin DMF düzeylerini belirlemek için okul türü değişkenine göre MH ve LR yöntemleriyle analizler R yazılımında difR paketi (Magis, Beland ve Raiche , 2016) ile gerçekleştirilmiştir. Maddelerin DMF düzeyleri ve hangi gruba avantaj sağladığı raporlaştırılmıştır.

Madde ve Test İstatistikleri

Bu bölümde, ABİDE 2016 uygulamasında kullanılan çoktan seçmeli maddelere ait 5600 öğrenci üzerinden hesaplanan madde istatistikleri ve öğrencilerin çoktan seçmeli maddelerden elde ettiği puanlara ilişkin betimsel istatistikler verilmiştir. Madde güçlükleri ve ayırt edicilik değerleri Tablo 6'da gösterilmiştir.

Tablo 6

Madde İstatistikleri

Madde No	Türkçe		Matematik		Fen Bilimleri	
	Ayırt		Ayırt		Ayırt	
	Güçlük (p)	Edicilik (r)	Güçlük (p)	Edicilik (r)	Güçlük (p)	Edicilik (r)
1	0,70	0,53	0,53	0,67	0,47	0,58
2	0,48	0,66	0,29	0,49	0,56	0,76
3	0,30	0,47	0,34	0,46	0,52	0,68
4	0,66	0,67	0,53	0,70	0,43	0,73
5	0,46	0,60	0,38	0,60	0,56	0,75
6	0,66	0,72	0,55	0,71	0,48	0,48
7	0,63	0,67	0,41	0,65	0,54	0,82
8	0,20	0,23	0,33	0,52	0,52	0,83
9	0,49	0,74	0,29	0,49	0,47	0,73

Tablo 6'ya göre Türkçe alt testindeki çoktan seçmeli maddelerin güçlük indeksi 0,20 ile 0,70 arasında değişmekte ve madde ayırt edicilik indeksi 0,23 ile 0,74 değişmektedir. Matematik alt testindeki çoktan seçmeli maddelerin madde güçlük indeksi 0,29 ile 0,55 arasında ve madde ayırt edicilik indeksi 0,46 ile 0,71 arasında değişmektedir. Fen bilimleri alt testindeki çoktan seçmeli maddelerin madde güçlük indeksi 0,43 ile 0,56 arasında ve madde ayırt edicilik indeksi 0,48 ile 0,83 arasında değişmektedir. Madde güçlük indeksi, sınava katılan adayların maddeye doğru yanıt verme oranı olarak tanımlanmıştır (Crocker ve Algina, 1986). Madde güçlük indeksi 0 ve 1 arasında değerler alır. Bu değer 1'e ne kadar yakınsa buna bağlı olarak maddenin o kadar kolay olduğu söylenebilir. Madde ayırt edicilik gücü indeksi, madde ile ölçülmek istenen özellik bakımından o özelliğe sahip olan bireylerle sahip olmayan bireyleri ayırt edebilme düzeyini göstermektedir (Turgut ve Baykul, 2015). Tekin'e (2010) göre madde ayırt edicilik indeksi 0,20'den yüksek olan maddeler kabul edilebilir düzeyde ayırt edicidir. Buna göre, ABİDE 2016 uygulamasında kullanılan alt test maddelerinin ayırt edici ve çoğunlukla orta güçlükte olduğu söylenebilir.

Öğrencilerin okul türü değişkeni bakımından ait oldukları gruba göre elde edilen madde güçlük indeksleri Tablo 7’de gösterilmiştir.

Tablo 7

Okul Türü Değişkenine Göre Alt Testlerin Madde Güçlük Düzeyleri

Madde No	Türkçe		Matematik		Fen Bilimleri	
	Devlet	Özel	Devlet	Özel	Devlet	Özel
1	0,71	0,76	0,51	0,69	0,48	0,61
2	0,47	0,69	0,28	0,47	0,55	0,77
3	0,28	0,47	0,34	0,50	0,50	0,74
4	0,66	0,79	0,51	0,72	0,42	0,67
5	0,45	0,63	0,38	0,53	0,52	0,75
6	0,65	0,82	0,54	0,74	0,45	0,57
7	0,65	0,78	0,38	0,64	0,51	0,82
8	0,21	0,23	0,32	0,46	0,49	0,80
9	0,48	0,75	0,30	0,45	0,46	0,69

Tablo 7’ye göre tüm alt testlerde özel okulda öğrenim gören öğrencilerin, devlet okulunda öğrenim gören öğrencilere göre daha yüksek başarı gösterdiği belirlenmiştir. Özel okulda öğrenim gören öğrencilerin tüm maddelerde maddeye doğru yanıt verme oranları, devlet okulunda öğrenim gören öğrencilerin maddeye doğru yanıt verme oranından yüksektir. Madde güçlük düzeyleri arasındaki en düşük fark 0,02; en yüksek fark 0,31 olarak gerçekleşmiştir. Okul türü değişkenine göre gruplara ait test istatistikleri Tablo 8’de sunulmuştur.

Tablo 8

Okul Türüne Göre Test İstatistikleri

Değişken			Devlet Okulu	Özel Okul
Alt Testler	Türkçe	\bar{x}	4,55	5,93
		ss	2,34	2,25
	Matematik	\bar{x}	3,56	5,19
		ss	2,20	2,65
	Fen Bilim.	\bar{x}	4,37	6,41
		ss	2,73	2,48

Tablo 8'e göre tüm alt testlerde özel okul öğrencileri daha başarılı olmuştur. Her iki grupta da en düşük başarı matematik alt testindedir.

Faktör Analizi

Bir maddenin DMF göstermesinin sebebi, test ile ölçülmek istenen özellikle ilişkili olmayan değişkenler sebebiyle testi alan farklı gruplardaki bireylerin, maddeye doğru yanıt verme olasılıklarının farklılaşmasından kaynaklanır. Bunun dışında maddelerin çok boyutluluğu da DMF'ye sebep olabilir. Maddelerin DMF göstermesinin sebebi test ile ölçülmek istenen tek faktörlü yapıyla ilişkili olmayan testi alan grupların özelliklerinden mi yoksa testin çok boyutlu olmasından mı kaynaklandığının belirlenmesi için tek boyutluluğun sınanması gerekir. DMF belirleme yöntemleri için tek boyutluluk yalnızca bir istatistiksel ön koşul değil, yanlılık çalışmaları için de esastır (Shepard, 1982). Bu sebeplerle DMF analizleri yapılmadan önce alt testlerin tek boyutluluk varsayımlarını karşılayıp karşılayamadığı incelenmiştir.

Tek boyutluluğun sınanması amacıyla FACTOR 10.8.4 programı ile tetrakorik korelasyona dayalı açımlayıcı faktör analizi uygulanmıştır. Faktör analizi öncesi verinin faktör analizine uygunluğunu belirlemek için Kaiser Meyer Olkin (KMO) katsayısı ve Bartlett küresellik testi değerleri incelenmiştir. Field (2000), KMO için en düşük değer 0,50 olması gerektiğini 0,50 altındaki değerler için verinin faktörleşemeyeceğini belirtmiştir. Ayrıca Tabachnick ve Fidell (2013), Bartlett testinin istatistiksel olarak anlamlı çıkmasının gerekli olduğunu belirtmiştir. Tablo 9'da KMO katsayısı ve Bartlett testine ilişkin değerler gösterilmiştir.

Tablo 9

KMO Katsayıları ve Bartlett Testi Değerleri

Testler	KMO	Bartlett Testi		
		χ^2	sd	p
Türkçe	0,867	7120,308	36	0,000
Matematik	0,834	6182,782	36	0,000
Fen Bilimleri	0,901	10812,246	36	0,000

Tablo 9'a göre bütün alt testler için Bartlett testi sonuçları istatistiksel olarak anlamlıdır ($p < 0,05$). Ayrıca KMO katsayıları ise 0,50'den büyüktür. Bunlara göre veri faktör analizine uygundur.

Tetrakorik korelasyona dayalı açıklayıcı faktör analizi sırasında boyut sayısının belirlenmesi için paralel analiz, özdeğerlere ait yamaç birikinti grafiği sonuçları kullanılmıştır. Paralel analizde, gerçek veriyle aynı sayıda değişken ve gözlem içeren bir veya daha fazla veri seti üretilir. Üretilen tesadüfi veri ile gerçek verinin açıklanan varyans yüzdeleri karşılaştırılır. Gerçek veriden elde edilen açıklanan varyans yüzdesinin, tesadüfi verinin açıklanan varyans yüzdesinden büyük olduğu durumlar dikkate alınarak faktör sayısı belirlenir (O'Connor, 2000). Tablo 10'da alt testlere ait paralel analiz tablosu sunulmuştur.

Tablo 10

Paralel Analiz Tablosu

Alt Testler	Faktör	Açıklanan Ortak Varyans Yüzdeleri		
		Gerçek Veri	Tesadüfi Veri	Tesadüfi Verinin 95.
Türkçe	F1	70,09	33,59	45,35
	F2	15,01	26,83	34,07
	F3	6,86	18,59	25,85
	F4	4,67	11,49	18,92
	F5	2,79	6,92	10,35
Matematik	F1	63,22	39,06	56,54
	F2	14,61	28,01	38,16
	F3	10,77	16,54	27,71
	F4	7,65	9,25	19,01
	F5	3,08	5,27	8,90
Fen Bilimleri	F1	73,59	39,42	56,73
	F2	13,39	28,07	38,80
	F3	5,83	16,32	28,10
	F4	3,79	9,16	20,82
	F5	2,21	5,12	8,12

Tablo 10'a göre tüm alt testlerde yalnızca birinci faktörlerde gerçek veriye ait açıklanan varyans yüzdeleri, tesadüfi veriye ait açıklanan varyans yüzdesinden büyüktür. Diğer faktörlerde ise gerçek veriye ait açıklanan varyans yüzdesinin

tesadüfi veriye ait açıklanan varyans yüzdesinden küçük olduğu görülmektedir. Bu sonuçlara göre alt testlerde tek boyutluluk sağlanmaktadır.

Alt testlerin boyut sayısının belirlenebilmesi için faktörlere ait özdeğerlerden ve yamaç birikinti grafiklerinden yararlanılmıştır. Tablo 11’de alt testlere ait özdeğerler ve açıklanan varyans yüzdeleri sunulmuştur.

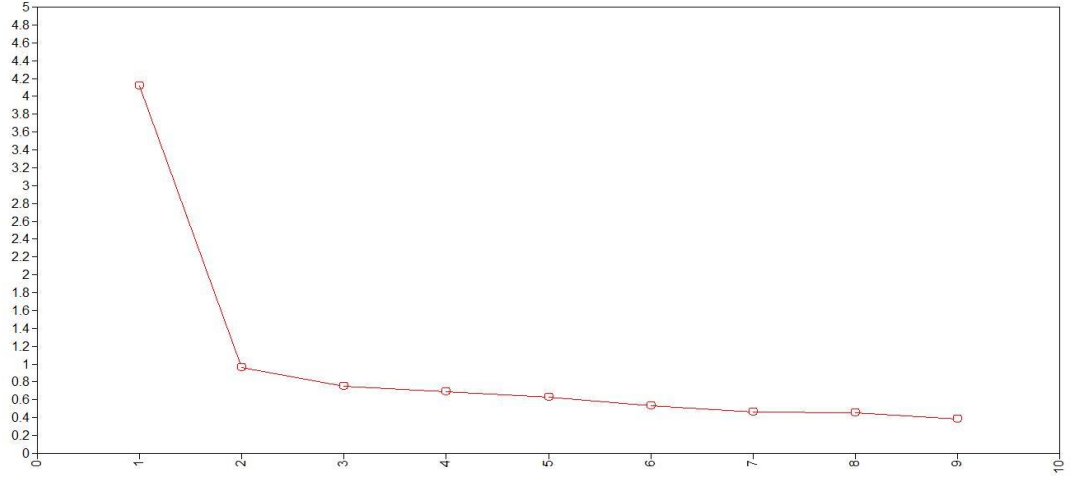
Tablo 11

Alt Testlere Ait Özdeğerler ve Açıklanan Varyans Yüzdeleri

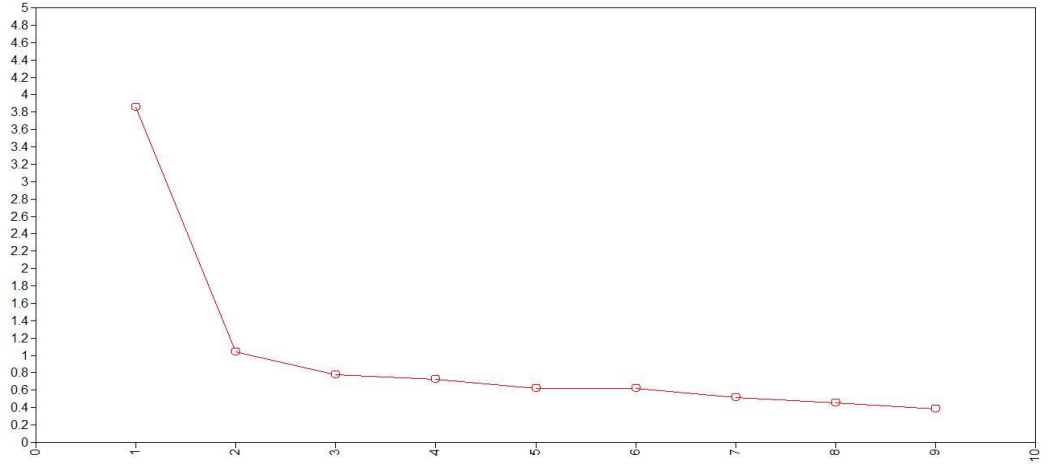
Testler	Faktör Sayısı	Özdeğer	Açıklanan Varyans (%)
Türkçe	1	4,121	43,43
Matematik	1	3,859	39,62
	2	1,041	11,47
Fen Bilimleri	1	4,815	52,28

Lord’a (1980) göre birinci faktörün özdeğeri ve açıkladığı varyans yüzdesi yüksek iken bu değerlerin ikinci faktörde düşmesi ve ikinci faktör ile sonraki faktörlerin özdeğerleri arasında yakınlık bulunması tek boyutluluğu gösterir. Bunun yanında özdeğerlere ait yamaç birikinti grafiğinde birinci faktörden sonra dikeyde ani bir düşüş sonrası diğer faktörlerde yatayda bir değişim gözleniyorsa tek boyutluluğa bir kanıt olarak sunulabilir (Çokluk, Şekercioğlu ve Büyüköztürk, 2016). Ayrıca Erkuş’a (2014) göre ilk faktörün özdeğeri sonraki faktörün 2-3 katından fazla ise yapı tek faktörlü olarak nitelendirilebilir.

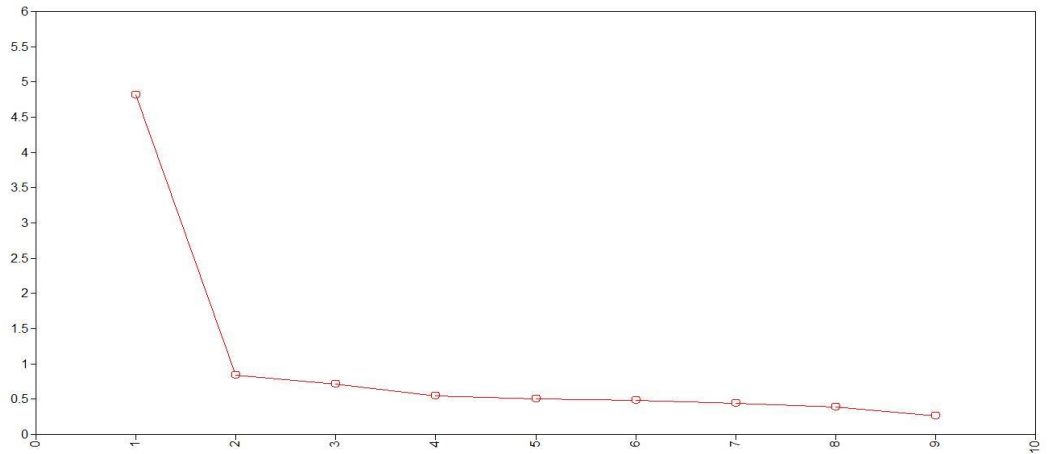
ABİDE çalışması 2016 uygulamasındaki Türkçe, matematik ve fen bilimleri alt testlerine ait yamaç birikinti (özdeğerlere ait çizgi grafiği) grafikleri Şekil 4, Şekil 5 ve Şekil 6’da sunulmuştur.



Şekil 4. Türkçe Alt Testine Ait Yamaç Birikinti Grafiği



Şekil 5. Matematik Alt Testine Ait Yamaç Birikinti Grafiği



Şekil 6. Fen Bilimleri Alt Testine Ait Yamaç Birikinti Grafiği

Grafiklere göre tüm alt testlerde birinci faktörden sonra ani bir düşüş olduğu görülmektedir. Paralel analiz sonuçları, faktörlere ait özdeğerler ve yamaç birikinti

grafikleri incelendiğinde Türkçe, matematik ve fen bilimleri alt testlerinin DMF analizlerinin tek boyutluluk varsayımını karşıladığı sonucuna ulaşılmıştır.

Eğilim Puanı Eşleştirme

EPE’de ilk adım, eğilim puanlarının hesaplanacağı ortak değişkenleri belirlemektir. Bu belirleme literatüre dayanmalıdır. Akademik başarının çıktı olarak ele alındığı eğitim araştırmalarında; demografik özellikler, gelir, aile eğitim düzeyi, aile yapısı uygun ortak değişkenler olarak ele alınır. Çıktı değişkenini açıklayabilen bu değişkenler, eğilim puanı oluşturmaya dahil edilmelidir (Fan ve Nowell, 2011). MEB (2017); sosyoekonomik düzey (SED), aile ilgisi, anne eğitim düzeyi değişkenlerinin öğrencilerin akademik başarısıyla ilişkili olduğunu belirlemiştir. Bu nedenlerle bu çalışmada SED, aile ilgisi, anne ve baba eğitim düzeyi ortak değişkenler olarak belirlenmiştir. İstatistiksel olarak gruplar arasında anlamlı farklılığın bulunduğu değişkenlerin ortak değişkenler olarak ele alınması gerekmektedir (Rosenbaum, 2002; Luellen, Shadish ve Clark, 2005). Bu çalışmada ortak değişken olarak belirlenen değişkenler için gerçekleştirilen bağımsız örneklem t testi sonuçları Tablo 12’de sunulmuştur.

Tablo 12

Eşleştirme Öncesi Ortak Değişkenlere İlişkin t Testi Sonuçları

Ortak	Okul Türü				t	sd	p	d
	Devlet		Özel					
	\bar{x}	ss	\bar{x}	ss				
SED	-0,07	0,93	1,47	0,99	-36,33	6328	0,000	1,60
Aile İlgisi	0,007	0,99	-0,14	0,97	3,35	6328	0,000	0,14
Anne E.	2,61	1,19	4,21	1,60	-28,67	6328	0,000	1,13
Baba E.	3,21	1,35	4,74	1,58	-24,72	6328	0,000	1,04

Tablo 12’de sunulan istatistikler incelendiğine SED, aile ilgisi, anne ve baba eğitim düzeyi değişkenlerinin okul türüne göre istatistiksel olarak anlamlı farklılık gösterdiği

görülmektedir. Bu sonuçlara göre bu değişkenler bakımından özel okulu ve devlet okulu öğrencileri arasında anlamlı farklılık olduğu ve buna bağlı olarak bu değişkenlerin öğrencilerin alt testlerde gösterdiği başarıda farklılık yaratabileceği söylenebilir. Seçilen bu dört ortak değişkenin (bağımsız değişken), okul türüyle (bağımlı değişken) ilişkisini istatistiksel olarak test etmek için lojistik regresyon analizi gerçekleştirilmiştir. Tablo 13'te analiz sonuçları gösterilmiştir.

Tablo 13

Ortak Değişkenlere İlişkin Lojistik Regresyon Analizi Sonuçları

Ortak Değişkenler	B	SH	Wald	sd	p	Odds Oranı	Odds Oranları için %95'lik Güven Aralığı	
							Alt Değer	Üst Değer
SED	1,476	0,090	267,762	1	0,000	4,375	3,666	5,221
Aile İlgisi	0,147	0,056	6,896	1	0,009	1,158	1,038	1,292
Anne E.	-0,008	0,049	0,029	1	0,864	0,992	0,901	1,091
Baba E.	-0,081	0,047	2,951	1	0,086	0,922	0,840	1,012

Bu dört ortak değişken ile öğrencilerin hangi okul türünde bulunduğu olasılığını belirlemek amacıyla bir model kurulmuştur ($\chi^2_{sd=4} = 996,93$; $p < 0,001$; Nagelkerke $R^2 = 0.330$). Ki-kare değerlerinin anlamlı olması kurulan bu modelin ortak değişkenlerin yordayıcılığı bağlamında iyi olduğunu göstermektedir.

Ortak değişkenlerin belirlenmesinden ve kontrolünden sonra EPE için eşleştirme yöntemi belirlenmelidir. Alanyazında farklı eşleştirme yöntemleri bulunmaktadır. Bu çalışmada en yakın komşu eşleştirme kullanılmıştır. En yakın komşu, EPE çalışmalarında en yaygın kullanılan eşleştirme yöntemidir (Thoemmes ve Kim, 2011). Bu yöntemde deney grubundaki birimler, eğilim puanlarının en yakın olduğu kontrol grubu birimleri ile eşleştirilir. Eşleştirme oranı 1:1, 1:2 ve 1:N şeklinde olabilir ancak bu çalışmada deney grubundaki (özel okul) her bir birimi kontrol grubundaki yalnızca bir birime eşleştiren 1:1 oranı kullanılmıştır. Eşleştirilen

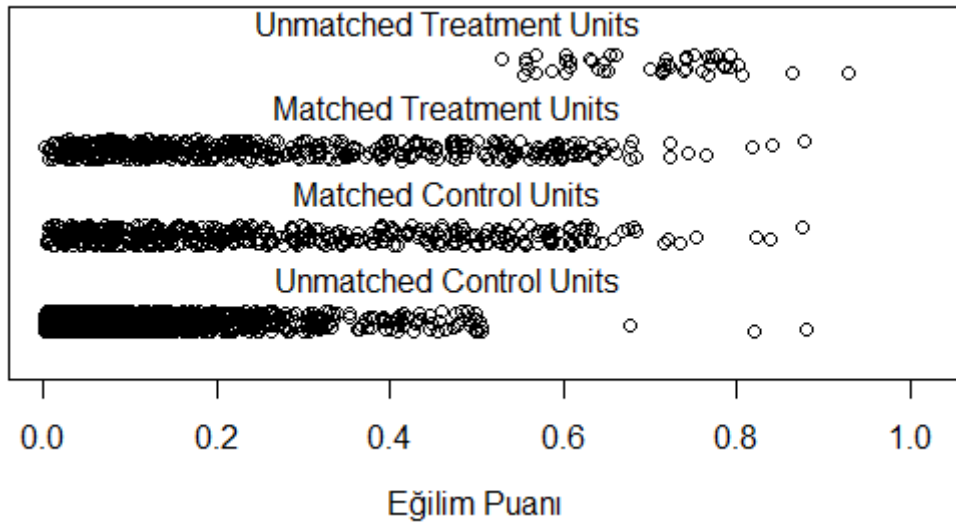
birimlerin eğilim puanları arasındaki farkın fazla olması kötü bir eşleştirmeye neden olabilir. Bu durumdan kaçınmak amacıyla eşleştirilecek birimlerin eğilim puanları arasındaki farkın maksimum değeri olan bir caliper tanımlanabilir. En yakın komşu eşleştirme yönteminin kullanıldığı çalışmalarda 0,10-0,25 aralığında caliper seçimleri yapılmıştır (Thoemmes ve Kim, 2011). Kötü eşleştirmelerden kaçınmak amacıyla bu çalışmada caliper “0.1” olarak belirlenmiştir. Eşleştirilecek birim mevcut olmadığında, eşleştirmenin kalitesini etkileyebileceği için birimler atılabilir (Stuart, 2010). Ortak değişkenler bakımından gruplar (özel okul-devlet okulu) arası dengenin sağlanabilmesi için eşleştirilmeyen birimlerin her iki gruptan da çıkarılması sağlanmıştır. Bu koşullara göre gerçekleştirilen EPE sonucu gruptaki kişi sayısı Tablo 14’te gösterilmiştir.

Tablo 14

Eşleştirme Sonrası Gruptaki Kişi Sayıları

	Devlet Okulu	Özel Okul
Toplam	5791	539
Eşleştirilen	489	489
Eşleştirilmeyen	5302	50
Atılan	0	0

EPE analizi sonrasında her iki gruptan 489 kişi birbiriyle eşleştirilmiş özel okuldan 50, devlet okulundan 5302 kişi bu eşleştirme dışında kalmıştır. Bu eşleştirmeye ilişkin jitter grafiği Şekil 7’de gösterilmiştir.



Şekil 7. Eğilim Puanı Eşleştirmeye İlişkin Jitter Grafiği

Şekil 7’de yer alan her bir daire bir öğrenciye karşılık gelmektedir. En üst satırda yer alan dağılım eşleştirilmeyen özel okul öğrencilerini, ikinci ve üçüncü satırlar sırasıyla eşleştirilen özel ve devlet okulu öğrencilerini, son satır ise eşleştirilmeyen devlet okulu öğrencilerini göstermektedir. İkinci ve üçüncü dağılımın benzer olması eşlemenin yakın olduğunu göstermektedir.

EPE sonrası ortak değişkenlerin kontrolünün uygun biçimde gerçekleştirilmesi ve kontrol analizlerinin sonuçlarını rapor etmek hayati öneme sahiptir (Zhang ve ark., 2019). Literatürde eşleştirme dengesini kontrol etmek için farklı stratejiler kullanılmıştır. İstatistiksel anlamlılık testleri (t testi, ki-kare), standartlaştırılmış ortalama farkları (SOF), grafik karşılaştırma en yaygın kullanıma sahip stratejilerdendir (Thoemmes ve Kim, 2011). Tablo 15’te eşleştirme sonrası ortak değişkenlere ilişkin bağımsız örneklem t testi analiz sonuçları yer almaktadır.

Tablo 15

Eşleştirme Sonrası Ortak Değişkenlere İlişkin T Testi Sonuçları

Ortak Değişkenler	Okul Türü				t	sd	p
	Devlet		Özel				
	\bar{x}	ss	\bar{x}	ss			
SED	1,306	0,98	1,323	0,92	-0,279	976	0,781
Aile İlgisi	-0,141	0,9	-0,141	0,96	-0,003	976	0,997
Anne E.	4,19	1,54	4,13	1,55	0,639	976	0,523
Baba E.	4,79	1,48	4,73	1,56	0,691	976	0,490

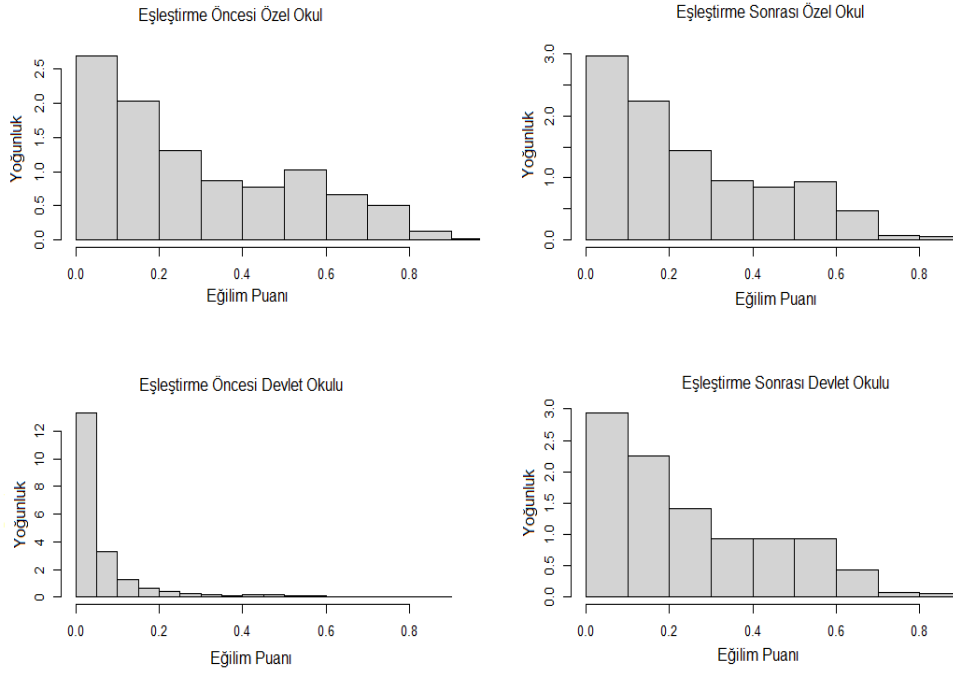
Tablo 15'e göre eşleştirme sonrası özel okul ve devlet okulu öğrencileri arasında ortak değişkenlerin hiçbirinde anlamlı farklılık olmadığı belirlenmiştir. Tablo 16'da ortak değişkenlere ilişkin eşleştirme öncesi ve sonrası SOF, eşleştirme sonrası varyans oranları sunulmuştur.

Tablo 16

Eşleştirme Öncesi ve Sonrası SOF Değerleri, Eşleştirme Sonrası Varyans Oranları

Ortak Değişkenler	SOF		Varyans Oranları
	Eşleştirme Öncesi	Eşleştirme Sonrası	
Eğilim Puanı	1,01	0,01	1,01
SED	1,59	0,01	0,87
Aile İlgisi	0,15	0,00	1,12
Anne E.	1,12	0,04	1,00
Baba E.	1,04	0,04	1,11

Rubin (2001), ortak deęişkenlerin dengesinin kontrolünde SOF deęerinin her bir deęişken için 0,25'ten az olması ve varyans oranlarının bire yakın olması gerektiğini belirtmiştir. Zhang ve ark. (2019), varyans oranlarının bir civarında olmasının iyi bir eşleştirmeye işaret ettiğini ve 2'den küçük varyans oranlarının kabul edilebilir olduğunu belirtmiştir. Tablo 16'da eşleştirme sonrası her bir deęişken için SOF deęerleri 0,25'ten küçük sifıra yakın, varyans oranları 2'den küçük ve bire yakın elde edilmiştir. Bu veriler eşleştirmenin başarılı olduğunu desteklemektedir. Şekil 8'de eşleştirme öncesi ve sonrası gruplara ait eğilim puanları dağılımı sunulmuştur.



Şekil 8. Eşleştirme Öncesi ve Sonrası Gruplara Ait Eğilim Puanı Dağılımları

Şekil 8'e göre eşleştirme öncesi gruplara ait eğilim puanları oldukça farklıyken eşleştirme sonrası dağılım çok benzer bir hale gelmiştir. Sayısal verilere ek olarak grafiklerde de iyi bir eşleştirme gerçekleştiği görülmektedir. Bu sonuçlara göre ortak deęişkenler bakımından özel okul ve devlet okulu grupları dengeli eşleşmiştir.

Bölüm 4

Bulgular ve Yorumlar

Bu bölümde araştırma kapsamında oluşturulan sorulara yönelik bulgu ve yorumlara yer verilmiştir. Bu amaçla ABİDE 2016 uygulamasında kullanılan Türkçe, matematik ve fen bilimleri alt testlerinde yer alan çoktan seçmeli maddelerin okul türü değişkenlerine göre EPE öncesi ve sonrası DMF gösterip göstermediği MH ve LR yöntemleriyle belirlenmeye çalışılmıştır.

Araştırma kapsamında elde edilen bulgular, araştırma sorularının sırasına göre yorumlanmıştır.

Alt Problem 1a'ya İlişkin Bulgular ve Yorumlar

ABİDE Türkçe alt testinde bulunan maddeler okul türüne göre DMF göstermekte midir?

ABİDE çalışması 2016 yılı uygulamasında Türkçe alt testindeki maddelerin okul türü değişkenine göre DMF gösterip göstermediğini belirlemek için MH yöntemi ile gerçekleştirilen analizlerden elde edilen sonuçlar Tablo 17'de sunulmuştur.

Tablo 17

Türkçe Alt Testine Ait Okul Türü Değişkenine Göre MH Analizi Sonuçları

Madde	α_{MH}	χ^2_{MH}	p	Δ_{MH}	DMF Düzeyi	Avantajlı Grup
1	0,739	4,846	0,027	0,710	A	Devlet
2	1,385	6,397	0,011	-0,766	A	Özel
3	0,891	0,900	0,342	0,270		
4	0,993	0,001	0,975	0,014		
5	1,063	0,224	0,635	-0,145		
6	1,057	0,065	0,798	-0,131		
7	0,963	0,040	0,840	0,087		
8	0,636	13,479	0,000	1,061	B	Devlet
9	1,738	16,214	0,000	-1,300	B	Özel

Tablo 17'de sunulan maddelerden sonuçları istatistiksel olarak anlamlı ($p < 0,05$) olanlarda $\Delta\alpha_{MH}$ değerleri incelenmiş ve MH yöntemi için önerilen Tablo 3'teki eşik değerleriyle kıyaslanarak maddenin hangi düzeyde DMF gösterdiği

belirlenmiştir. Ayrıca $\Delta\alpha_{MH}$ değerinin negatif ya da pozitif olmasına göre hangi grubun avantajlı olduğu kararı verilmiştir.

Tablo 17'e göre Türkçe alt testinde iki madde A, iki madde ise B düzeyinde DMF içermektedir, C düzeyinde DMF içeren madde gözlenmemiştir. Bu sonuç bize MH yöntemi ile gerçekleştirilen analizlerde okul türü değişkenine göre Türkçe alt testinde 8 numaralı maddenin devlet okulunda öğrenim gören öğrenciler lehine, 9 numaralı maddenin ise özel okulda öğrenim gören öğrenciler lehine orta düzeyde DMF içerdiğini göstermiştir.

ABİDE çalışması 2016 yılı uygulamasında Türkçe alt testindeki maddelerin okul türü değişkenine göre DMF gösterip göstermediğini belirlemek için LR yöntemi ile gerçekleştirilen analizlerden elde edilen sonuçlar Tablo 18'de sunulmuştur.

Tablo 18

Türkçe Alt Testine Ait Okul Türü Değişkenine Göre LR Analizi Sonuçları

Madde	ΔR^2	$\Delta\chi^2$	DMF Düzeyi	DMF Biçimi	Avantajlı Grup
1	0,790	15,806	C	TB	Devlet
2	0,000	14,376	A	TB	Özel
3	0,000	10,889	A	TBO	
4	0,000	5,212			
5	0,000	7,014	A	TBO	
6	0,000	0,259			
7	0,000	1,045			
8	0,002	18,081	A	TB	Devlet
9	0,000	18,360	A	TB	Özel

Tablo 18'de sunulan maddelerden, $\Delta\chi^2$ değerleri istatistiksel olarak anlamlı ($p<0,05$) olanlar incelenmiş ve DMF düzeyini belirlemek amacıyla bu maddelerin ΔR^2 değerleri, Tablo 4'teki Bakan Kalaycıoğlu ve Kelecioğlu (2011) tarafından önerilen eşik değerleri ile karşılaştırılmıştır. Tablo 18'e göre LR yöntemi ile gerçekleştirilen analiz sonuçlarında Türkçe alt testinde beş madde okul türü değişkenine göre A düzeyinde, bir madde C düzeyinde DMF içermektedir. Bu maddelerden 1 numaralı madde devlet okulunda öğrenim gören öğrenciler lehine yüksek düzeyde TB DMF gösterirken 2 ve 9 numaralı maddeler özel okulda öğrenim gören öğrenciler lehine, 8 numaralı madde devlet okulda öğrenim gören öğrenciler

lehine ihmal edilebilir düzeyde TB DMF göstermiştir. Ayrıca 3 ve 5 numaralı maddelerin ihmal edilebilir düzeyde TBO DMF gösterdiği belirlenmiştir.

Alt Problem 1b'ye İlişkin Bulgular ve Yorumlar

ABİDE matematik alt testinde bulunan maddeler okul türüne göre DMF göstermekte midir?

ABİDE çalışması 2016 yılı uygulamasında matematik alt testindeki maddelerin okul türü değişkenine göre DMF gösterip göstermediğini belirlemek için MH yöntemi ile gerçekleştirilen analizlerden elde edilen sonuçlar Tablo 19'da sunulmuştur.

Tablo 19

Matematik Alt Testine Ait Okul Türü Değişkenine Göre MH Analizi Sonuçları

Madde	α_{MH}	χ^2_{MH}	p	Δ_{MH}	DMF Düzeyi	Avantajlı Grup
1	1,553	4,602	0,031	-1,034	B	Özel
2	1,043	0,089	0,764	-0,099		
3	1,183	2,015	0,155	-0,395		
4	1,074	0,190	0,662	-0,168		
5	0,816	2,904	0,088	0,475		
6	1,154	0,881	0,347	-0,338		
7	1,439	7,448	0,006	-0,855	A	Özel
8	0,687	10,286	0,001	0,880	A	Devlet
9	0,774	4,099	0,042	0,601	A	Devlet

Tablo 19'da sunulan maddelerden sonuçları istatistiksel olarak anlamlı ($p < 0,05$) olanlarda $\Delta\alpha_{MH}$ değerleri incelenmiş ve MH yöntemi için önerilen Tablo 3'teki eşik değerleriyle kıyaslanarak maddenin hangi düzeyde DMF gösterdiği belirlenmiştir.

Tablo 19'a göre bir madde B düzeyinde DMF gösterirken, üç madde A düzeyinde DMF göstermiştir. Bu sonuçlar bize okul türü değişkenine göre matematik alt testinde 1 numaralı maddenin özel okulda öğrenim gören öğrenciler lehine orta düzeyde DMF içerdiğini göstermiştir.

ABİDE çalışması 2016 yılı uygulamasında matematik alt testindeki maddelerin okul türü değişkenine göre DMF gösterip göstermediğini belirlemek için

LR yöntemi ile gerçekleştirilen analizlerden elde edilen sonuçlar Tablo 20’de sunulmuştur.

Tablo 20

Matematik Alt Testine Ait Okul Türü Değişkenine Göre LR Analizi Sonuçları

Madde	ΔR^2	$\Delta \chi^2$	DMF Düzeyi	DMF Biçimi	Avantajlı Grup
1	0,000	1,910			
2	0,000	2,032			
3	0,000	2,336			
4	0,000	4,195			
5	0,000	5,564			
6	0,000	0,006			
7	0,000	20,084	A	TB	Özel
8	0,000	17,240	A	TB	Devlet
9	0,000	19,418	A	TBO	

Tablo 4.12.’de sunulan maddelerden, $\Delta \chi^2$ değerleri istatistiksel olarak anlamlı ($p < 0,05$) olanlar incelenmiş ve DMF düzeyini belirlemek amacıyla bu maddelerin ΔR^2 değerleri, Tablo 20’deki Bakan Kalaycıoğlu ve Kelecioğlu (2011) tarafından önerilen eşik değerleri ile karşılaştırılmıştır.

Tablo 20’ye göre LR yöntemi ile gerçekleştirilen analiz sonuçlarında matematik alt testinde üç madde okul türü değişkenine göre ihmal edilebilir düzeyde DMF içermektedir.

Alt Problem 1c’ye İlişkin Bulgular ve Yorumlar

ABİDE fen bilimleri alt testinde bulunan maddeler okul türüne göre DMF göstermekte midir?

ABİDE çalışması 2016 yılı uygulamasında fen bilimleri alt testindeki maddelerin cinsiyet değişkenine göre DMF gösterip göstermediğini belirlemek için MH yöntemi ile gerçekleştirilen analizlerden elde edilen sonuçlar Tablo 21’de sunulmuştur.

Tablo 21

Fen Bilimleri Alt Testine Ait Okul Türü Değişkenine Göre MH Analizi Sonuçları

Madde	α_{MH}	χ^2_{MH}	p	Δ_{MH}	DMF Düzeyi	Avantajlı Grup
1	0,705	9,555	0,002	0,820	A	Devlet
2	0,826	1,496	0,221	0,448		
3	1,274	3,169	0,075	-0,570		
4	1,220	2,222	0,136	-0,467		
5	0,872	0,857	0,354	0,321		
6	0,750	7,024	0,008	0,673	A	Devlet
7	1,657	8,352	0,003	-1,186	B	Özel
8	1,296	2,223	0,135	-0,609		
9	1,215	2,034	0,153	-0,458		

Tablo 21’de sunulan maddelerden sonuçları istatistiksel olarak anlamlı ($p < 0,05$) olanlarda $\Delta\alpha_{MH}$ değerleri incelenmiş ve MH yöntemi için önerilen Tablo ’teki eşik değerleriyle kıyaslanarak maddenin hangi düzeyde DMF gösterdiği belirlenmiştir.

Tablo 21’e göre yalnızca iki madde A düzeyinde DMF gösterirken, bir madde B düzeyinde DMF göstermiştir. Bu sonuçlar bize okul türü değişkenine göre fen bilimleri alt testinde 1 ve 6 numaralı maddelerin devlet okulunda öğrenim gören öğrenciler lehine ihmal edilebilir düzeyde DMF içerdiğini, 7 numaralı maddenin ise özel okulda öğrenim gören öğrenciler lehine orta düzeyde DMF içerdiğini göstermiştir.

ABİDE çalışması 2016 yılı uygulamasında fen bilimleri alt testindeki maddelerin okul türü değişkenine göre DMF gösterip göstermediğini belirlemek için LR yöntemi ile gerçekleştirilen analizlerden elde edilen sonuçlar Tablo 22’de sunulmuştur.

Tablo 22

Fen Bilimleri Alt Testine Ait Okul Türü Değişkenine Göre LR Analizi Sonuçları

Madde	ΔR^2	$\Delta \chi^2$	DMF Düzeyi	DMF Biçimi	Avantajlı Grup
1	0,000	14,335	A	TB	Devlet
2	0,000	6,814	A	TB	Devlet
3	0,000	7,269	A	TB	Özel
4	0,000	14,733	A	TBO	
5	0,000	2,837			
6	0,002	16,971	A	TB	Devlet
7	0,000	13,962	A	TB	Özel
8	0,000	5,540			Devlet
9	0,000	7,767	A	TBO	

Tablo 22’de sunulan maddelerden, $\Delta \chi^2$ değerleri istatistiksel olarak anlamlı ($p < 0,05$) olanlar incelenmiş ve DMF düzeyini belirlemek amacıyla bu maddelerin ΔR^2 değerleri, Tablo 3’teki Bakan Kalaycıoğlu ve Kelecioğlu (2011) tarafından önerilen eşik değerleri ile karşılaştırılmıştır. Tablo 22’ye göre yedi madde ihmal edilebilir düzeyde DMF göstermiştir.

Alt Problem 2a’ya İlişkin Bulgular ve Yorumlar

EPE sonrasında ABİDE Türkçe alt testinde bulunan maddeler okul türüne göre DMF göstermekte midir?

ABİDE çalışması 2016 yılı uygulamasında Türkçe alt testindeki maddelerin EPE sonrasında okul türü değişkenine göre DMF gösterip göstermediğini belirlemek için MH yöntemi ile gerçekleştirilen analizlerden elde edilen sonuçlar Tablo 23’te sunulmuştur.

Tablo 23

Türkçe Alt Testine Ait Okul Türü Değişkenine Göre MH Analizi Sonuçları

Madde	α_{MH}	χ^2_{MH}	p	Δ_{MH}	DMF Düzeyi	Avantajlı Grup
1	0,807	1,114	0,291	0,502		
2	1,371	3,150	0,075	-0,742		
3	1,005	0,002	0,961	-0,012		
4	1,036	0,006	0,935	-0,084		
5	0,997	0,003	0,950	0,006		
6	1,204	0,546	0,459	-0,437		

7	0,977	0,000	0,985	0,053		
8	0,685	4,661	0,030	0,886	A	Devlet
9	1,182	0,584	0,444	-0,394		

Tablo 23'te sunulan maddelerden sonuçları istatistiksel olarak anlamlı ($p < 0,05$) olanlarda $\Delta\alpha_{MH}$ değerleri incelenmiş ve MH yöntemi için önerilen Tablo 3'teki eşik değerleriyle kıyaslanarak maddenin hangi düzeyde DMF gösterdiği belirlenmiştir.

Tablo 23'e göre Türkçe alt testinde bir madde A düzeyinde DMF içermektedir, B ve C düzeyinde DMF içeren madde gözlenmemiştir. Bu sonuç bize MH yöntemi ile gerçekleştirilen analizlerde okul türü değişkenine göre Türkçe alt testinde 8 numaralı maddenin devlet okulunda öğrenim gören öğrenciler lehine ihmal edilebilir düzeyde DMF içerdiğini göstermiştir.

ABİDE çalışması 2016 yılı uygulamasında Türkçe alt testindeki maddelerin EPE sonrasında okul türü değişkenine göre DMF gösterip göstermediğini belirlemek için LR yöntemi ile gerçekleştirilen analizlerden elde edilen sonuçlar Tablo 24'te sunulmuştur.

Tablo 24

Türkçe Alt Testine Ait Okul Türü Değişkenine Göre LR Analizi Sonuçları

Madde	ΔR^2	$\Delta\chi^2$	DMF Düzeyi	DMF Biçimi	Avantajlı Grup
1	0,004	4,021			
2	0,003	3,812			
3	0,000	0,776			
4	0,005	5,089			
5	0,000	0,564			
6	0,001	0,867			
7	0,000	0,201			
8	0,007	5,856			
9	0,001	1,468			

Tablo 24'te sunulan maddelerden, $\Delta\chi^2$ değerleri istatistiksel olarak anlamlı ($p < 0,05$) olanlar incelenmiş ve DMF düzeyini belirlemek amacıyla bu maddelerin ΔR^2 değerleri, Tablo 4'teki Bakan Kalaycıoğlu ve Kelecioğlu (2011) tarafından önerilen eşik değerleri ile karşılaştırılmıştır. Tablo 24'e göre LR yöntemi ile

gerçekleştirilen analiz sonuçlarında Türkçe alt testinde DMF içeren madde bulunmamaktadır.

Alt Problem 2b'ye İlişkin Bulgular ve Yorumlar

EPE sonrasında ABİDE matematik alt testinde bulunan maddeler okul türüne göre DMF göstermekte midir?

ABİDE çalışması 2016 yılı uygulamasında matematik alt testindeki maddelerin EPE sonrasında okul türü değişkenine göre DMF gösterip göstermediğini belirlemek için MH yöntemi ile gerçekleştirilen analizlerden elde edilen sonuçlar Tablo 25'te sunulmuştur.

Tablo 25

Matematik Alt Testine Ait Okul Türü Değişkenine Göre MH Analizi Sonuçları

Madde	α_{MH}	χ^2_{MH}	p	Δ_{MH}	DMF Düzeyi	Avantajlı Grup
1	1,338	0,998	0,317	-0,684		
2	0,844	0,923	0,336	0,396		
3	1,034	0,016	0,896	-0,078		
4	1,225	0,988	0,320	-0,478		
5	0,977	0,003	0,950	0,052		
6	1,431	3,279	0,070	-0,842		
7	1,188	0,826	0,363	-0,408		
8	0,702	4,439	0,035	0,828	A	Devlet
9	0,833	0,978	0,322	0,427		

Tablo 19'da sunulan maddelerden sonuçları istatistiksel olarak anlamlı ($p < 0,05$) olanlarda $\Delta\alpha_{MH}$ değerleri incelenmiş ve MH yöntemi için önerilen Tablo 3'teki eşik değerleriyle kıyaslanarak maddenin hangi düzeyde DMF gösterdiği belirlenmiştir.

Tablo 25'e göre bir madde A düzeyinde DMF gösterirken B ve C düzeyinde DMF gösteren madde bulunmamaktadır. Bu sonuçlar bize okul türü değişkenine göre matematik alt testinde 8 numaralı maddenin devlet okulunda öğrenim gören öğrenciler lehine ihmal edilebilir düzeyde DMF içerdiğini göstermiştir.

ABİDE çalışması 2016 yılı uygulamasında matematik alt testindeki maddelerin EPE sonrasında okul türü değişkenine göre DMF gösterip

göstermediğini belirlemek için LR yöntemi ile gerçekleştirilen analizlerden elde edilen sonuçlar Tablo 26’da sunulmuştur.

Tablo 26

Matematik Alt Testine Ait Okul Türü Değişkenine Göre LR Analizi Sonuçları

Madde	ΔR^2	$\Delta \chi^2$	DMF Düzeyi	DMF Biçimi	Avantajlı Grup
1	0,000	0,174			
2	0,004	5,204			
3	0,000	0,110			
4	0,003	3,732			
5	0,000	0,352			
6	0,002	2,399			
7	0,004	5,620			
8	0,004	5,056			
9	0,005	5,933			

Tablo 26’da sunulan maddelerden, $\Delta \chi^2$ değerleri istatistiksel olarak anlamlı ($p < 0,05$) olanlar incelenmiş ve DMF düzeyini belirlemek amacıyla bu maddelerin ΔR^2 değerleri, Tablo 20’deki Bakan Kalaycıoğlu ve Kelecioğlu (2011) tarafından önerilen eşik değerleri ile karşılaştırılmıştır.

Tablo 26’ya göre LR yöntemi ile gerçekleştirilen analiz sonuçlarında matematik alt testinde DMF içeren madde bulunmamaktadır.

Alt Problem 2c’ye İlişkin Bulgular ve Yorumlar

EPE sonrasında ABİDE fen bilimleri alt testinde bulunan maddeler okul türüne göre DMF göstermekte midir?

ABİDE çalışması 2016 yılı uygulamasında fen bilimleri alt testindeki maddelerin EPE sonrasında okul türü değişkenine göre DMF gösterip göstermediğini belirlemek için MH yöntemi ile gerçekleştirilen analizlerden elde edilen sonuçlar Tablo 27’de sunulmuştur.

Tablo 27

Fen Bilimleri Alt Testine Ait Okul Türü Değişkenine Göre MH Analizi Sonuçları

Madde	α_{MH}	χ^2_{MH}	p	Δ_{MH}	DMF Düzeyi	Avantajlı Grup
1	0,737	3,425	0,064	0,716		
2	0,805	0,892	0,344	0,509		
3	1,767	9,961	0,001	-1,338	B	Özel
4	1,074	0,106	0,744	-0,169		
5	0,739	2,108	0,146	0,710		
6	0,898	0,426	0,513	0,252		
7	1,552	3,277	0,070	-1,033		
8	0,851	0,340	0,559	0,379		
9	1,144	0,435	0,509	-0,317		

Tablo 21’de sunulan maddelerden sonuçları istatistiksel olarak anlamlı ($p < 0,05$) olanlarda $\Delta\alpha_{MH}$ değerleri incelenmiş ve MH yöntemi için önerilen Tablo 21’teki eşik değerleriyle kıyaslanarak maddenin hangi düzeyde DMF gösterdiği belirlenmiştir.

Tablo 27’ye göre yalnızca bir madde B düzeyinde DMF gösterirken, bir A ve C düzeyinde DMF gösteren madde bulunmamaktadır. Bu sonuçlar bize okul türü değişkenine göre fen bilimleri alt testinde 3 numaralı maddenin özel okulunda öğrenim gören öğrenciler lehine orta düzeyde DMF içerdiğini göstermiştir.

ABİDE çalışması 2016 yılı uygulamasında fen bilimleri alt testindeki maddelerin EPE sonrasında okul türü değişkenine göre DMF gösterip göstermediğini belirlemek için LR yöntemi ile gerçekleştirilen analizlerden elde edilen sonuçlar Tablo 28’de sunulmuştur.

Tablo 28

Fen Bilimleri Alt Testine Ait Okul Türü Değişkenine Göre LR Analizi Sonuçları

Madde	ΔR^2	$\Delta \chi^2$	DMF Düzeyi	DMF Biçimi	Avantajlı Grup
1	0,007	7,209	A	TB	Devlet
2	0,001	1,979			
3	0,007	8,444	A	TB	Özel
4	0,004	5,416			
5	0,003	4,023			
6	0,004	4,273			

7	0,003	4,383
8	0,001	1,604
9	0,000	0,725

Tablo 28’de sunulan maddelerden, $\Delta\chi^2$ değerleri istatistiksel olarak anlamlı ($p<0,05$) olanlar incelenmiş ve DMF düzeyini belirlemek amacıyla bu ΔR^2 değerleri, Tablo 3’teki Bakan Kalaycıoğlu ve Kelecioğlu (2011) tarafından önerilen eşik değerleri ile karşılaştırılmıştır.

Tablo 28’e göre iki madde A düzeyde DMF göstermiştir. Bu maddelerden biri devlet diğeri ise özel okulda öğrenim gören öğrenciler lehine ihmal edilebilir düzeyde DMF göstermiştir.

Alt Problem 3’e İlişkin Bulgular ve Yorumlar

Eğilim puanları eşleştirme öncesi ve sonrası gerçekleştirilen DMF analizi sonuçları arasında farklılık bulunmakta mıdır?

ABİDE çalışması 2016 yılı uygulamasında alt testlerdeki maddelerin EPE öncesi ve sonrasında okul türü değişkenine göre DMF gösterip göstermediğini belirlemek için MH ve LR yöntemleriyle gerçekleştirilen analizlerden elde edilen sonuçlara göre hangi maddenin ne düzeyde DMF gösterdiği Tablo 29’da sunulmuştur.

Tablo 29

EPE Öncesi ve Sonrası DMF Analizi Sonuçları

Madde	EPE Öncesi						EPE Sonrası					
	Türkçe		Matematik		Fen Bilim.		Türkçe		Matematik		Fen Bilim.	
	MH	LR	MH	LR	MH	LR	MH	LR	MH	LR	MH	LR
1	A	C	B		A	A						A
2	A	A				A						
3		A				A					B	A
4						A						
5		A										
6					A	A						
7			A	A	B	A						
8	B	A	A	A			A		A			
9	B	A	A	A		A						

Tablo 29'a göre EPE sonrası gerekleřtirilen DMF analizlerinde DMF gsteren madde sayısında dikkat ekici bir dřř olmuřtur. EPE ncesi MH yntemi ile gerekleřtirilen DMF analizleri sonucunda yedi madde A, drt madde B dzeyinde DMF gstermiřtir. EPE ncesi LR yntemi ile gerekleřtirilen DMF analizleri sonucunda 15 madde A, bir madde ise C dzeyinde DMF gstermiřtir. EPE sonrası; MH yntemi ile gerekleřtirilen DMF analizleri sonucunda iki madde A, bir madde B dzeyinde DMF gsterirken LR ile gerekleřtirilen DMF analizleri sonucunda iki madde A dzeyinde DMF gstermiřtir. Bu sonular EPE analizleri sonrası DMF'li olarak iřaretlenen madde sayısında nemli azalma olduėunu ortaya koymuřtur.

Bölüm 5

Sonuç, Tartışma ve Öneriler

Bu bölümde araştırmadan elde edilen bulgulara dayalı sonuçlar ve öneriler yer almaktadır.

Sonuç

Çeşitli özellikleri bakımından grupların karşılaştırıldığı gözlemsel eğitim araştırmalarında gruplara atama üzerinde araştırmacının kontrolü olmayabilir. Bu nedenle gruplar, bazı ortak değişkenler bakımından sistematik olarak farklılık gösterebilir (Rosenbaum ve Rubin, 1983). Bu farklılık grupların karşılaştırılmasında yanlılığa ve karşılaştırmaya dayalı tahminlerin hatalı olmasına neden olabilir (Wu ve ark, 2017). Bu durumun olumsuz etkilerinden kaçınmak için grupların bu değişkenler bakımında eşleştirilmesinde Rosenbaum ve Rubin (1983) tarafından eğilim puanı eşleştirilmesi (EPE) model olarak önerilmiştir. Grupların karşılaştırıldığı DMF çalışmalarında ortak değişkenler bakımında grup farklılıklarını ortadan kaldırmak; DMF kaynaklarını saflaştıracak ve DMF'nin nedensel etkisinin tahmin edilmesine katkı sağlayacaktır (Liu ve ark., 2016).

EPE öncesi okul türüne göre gerçekleştirilen DMF analizleri sonucunda; MH yöntemine göre dört maddenin orta düzeyde DMF gösterdiği ve LR yöntemine göre bir maddenin yüksek düzeyde DMF gösterdiği belirlenmiştir. MH yöntemi ile orta düzeyde DMF gösterdiği belirlenen dört maddeden üçü özel okulda öğrenim gören öğrencilere avantaj sağlarken bir maddenin ise devlet okulunda öğrenim gören öğrencilere avantaj sağladığı görülmüştür. LR yöntemi ile yüksek düzeyde DMF gösterdiği belirlenen bir madde devlet okulunda öğrenim gören öğrencilere avantaj sağlamıştır. Ayrıca MH yöntemi ile yedi, LR yöntemi ile 15 maddenin önemsenmeyecek düzeyde DMF gösterdiği belirlenmiştir.

EPE sonrası okul türüne göre gerçekleştirilen DMF analizleri sonucunda; MH yöntemine göre bir maddenin orta düzeyde DMF gösterdiği belirlenirken LR yöntemiyle gerçekleştirilen analizler sonucunda orta ve yüksek düzey DMF gösteren madde bulunmamaktadır. MH yöntemi ile DMF gösterdiği belirlenen bir maddenin, özel okulda öğrenim gören öğrencilere avantaj sağladığı görülmüştür. Ayrıca MH ve LR yöntemleriyle gerçekleştirilen analizlerde ikişer maddenin önemsenmeyecek düzeyde DMF gösterdiği belirlenmiştir.

EPE öncesi ve sonrası sonuçlar karşılaştırıldığında her düzeyde DMF'li olarak işaretlenen madde sayısında dikkate değer bir düşüş olmuştur. Önemsenecek düzeyde DMF gösteren madde sayısı MH yöntemi ile gerçekleştirilen analizlerde yediden ikiye, LR yöntemiyle gerçekleştirilen analizlerde 15'ten ikiye düşmüştür. Eşleştirme sonrası orta ve yüksek düzeyde DMF gösteren madde sayısı yalnızca birdir.

Tartışma

Bu araştırmada EPE öncesi ve sonrası okul türüne göre DMF gösteren maddelerin olup olmadığı belirlenmiş ve DMF miktarlarındaki değişim incelenmiştir. Elde edilen sonuçlara göre ortak değişkenler bakımından grupların dengelenmesiyle DMF'li olarak işaretlenen madde sayısında azalma olacağı söylenebilir. Wu ve Ercikan (2006), grupların ortak değişkenler bakımından dengelenmesinin DMF kaynakları üzerinde kontrol sağladığını belirtmiştir. Ayrıca bu sonuçlar EPE'ye dayalı DMF çalışmalarının sonuçlarıyla da benzerlik göstermektedir (Lee ve Geisinger, 2014; Liu ve ark., 2016; Arıkan, Vijver ve Yağmur, 2018). DMF'li madde sayısının azalmasıyla seçilen ortak değişkenler bakımından DMF kaynaklarının kontrolünün sağlandığı söylenebilir. Liu ve ark. (2016), EPE ile DMF'ye neden olabilecek ortak değişkenler üzerinde kontrol sağlandığı sonucuna ulaşmıştır.

DMF çalışmalarında önemli sorunlardan biri DMF'nin kaynaklarını tanımlamaktır (Angoff, 1993). Ortak değişken olarak belirlenen SED, aile ilgisi, anne ve baba eğitim düzeyinin öğrencilerin test performansıyla ilişkili değişkenlerdir (Fan ve Nowell, 2011; Lee ve Geisinger, 2014; Arıkan, Vijver ve Yağmur, 2018; OECD, 2019). Bu değişkenler bakımından dengeli gruplar oluşturarak gerçekleştirilen DMF analizleri sonucunda DMF'li olarak işaretlenen madde sayısında önemli düşüş olmasıyla bu değişkenlerin okul türü için potansiyel bir DMF kaynağı olarak görülebilir. Ayrıca DMF analizleri sonrası başlatılacak yargısal bir süreç olan yanlılık çalışmalarında analiz sonuçlarının daha doğru yorumlanabileceği, eşleştirme öncesi ve sonrası gerçekleştirilen DMF analizleri sonuçları arasındaki farklılığın bu yargısal sürece dayanak oluşturacağı söylenebilir. Liu ve ark. (2016), EPE'ye dayalı DMF yaklaşımının, araştırmacılara ve test geliştiricisine nedensel DMF'yi incelemek için daha kesin bir araç sağlayabileceğini ve olası yanlı maddeleri düzeltme veya

kaldırma konusunda daha doğru bir karar vermeye yardımcı olabileceğini belirlemişlerdir.

Araştırma sonuçları bağlamında DMF'nin doğrudan okul türüne atfedilmesinin hatalı olabileceği, DMF'nin ortak değişkenler olarak belirlenen bu karıştırıcı değişkenlerden kaynaklandığı söylenebilir. Yine bu bağlamda EPE ile eğitim çalışmalarında neden-sonuç ilişkisini kuvvetlendiği ve yanlış kararlar alma olasılığının azaldığı söylenebilir.

Öneriler

Ulusal ve uluslararası test geliştirme çalışmaları yürüten tüm kuruluşlar tarafından tüm sınavlara yönelik EPE'ye dayalı DMF çalışmaları gerçekleştirilebilir ve bu çalışmalardan elde edilecek sonuçlar yardımıyla DMF kaynakları daha somut biçimde ortaya koyulabilir. Bu çalışmaların sonuçları test geliştirme süreçlerine yansıtılabilir, yanlışlık kaynağı olabilecek maddeler üzerinde kontrol sağlanabilir.

Bu çalışma kapsamında MH ve LR yöntemleriyle DMF analizleri gerçekleştirilmiştir. Farklı yöntemlerin DMF belirlemedeki gücünün; odak ve referans grup örneklem büyüklüğü, odak ve referans grup arasındaki yetenek farkı, testteki madde sayısı gibi değişkenlerden etkilendiği görülmektedir. Ayrıca yöntemlerin 1. Tip hatayı kontrol güçleri farklı koşullar altında değişim göstermektedir. Araştırmalarda bu değişkenler göz önünde bulundurularak daha iyi sonuç veren diğer DMF belirleme yöntemleri kullanılabilir.

Bu çalışma kapsamında DMF analizleri okul türü değişkenine göre gerçekleştirilmiştir. DMF analizleri, yanlışlık yaratabilecek yerleşim yeri, kültür, dil, dezavantajlı olma durumu gibi farklı değişkenlere göre gerçekleştirilebilir.

Bu çalışmada EPE analizlerinde kullanılan ortak değişkenler, ABİDE 2016 anket verilerinden elde edilen değişkenlerle sınırlıdır. Bu sebeplerle SED, aile ilgisi, anne ve baba eğitim düzeyi ortak değişkenler olarak belirlenmiştir. Araştırmalarda DMF kaynağı olabilecek teknolojik imkanlar, öğrenme fırsatları, eğitim harcamaları, öğrenmeye ayrılan zaman, ilgi, değer gibi diğer değişkenler de ortak değişken olarak eşleştirme algoritmasına eklenebilir.

EPE analizleri en yakın komşuluk eşleştirme yöntemi ile gerçekleştirilmiştir. Araştırmalarda tam eşleştirme, optimal eşleştirme gibi diğer eşleştirme yöntemlerinin kullanıldığı EPE analizleri gerçekleştirilebilir.

Bu çalışma kapsamında EPE öncesi ve sonrası DMF incelemesi gerçekleştirilmiştir. Grupların karşılaştırılmasına dayalı farklı eğitim araştırmalarında nedensel çıkarımların geçerliğini arttırmak amacıyla EPE'ye dayalı incelemeler gerçekleştirilebilir.

Kaynaklar

- American Educational Research Association., American Psychological Association., National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*.
- Angoff, H. W. (1993). Perspectives on Differential Item Functioning Methodology. P.W. Holland ve H. Wainer (Ed.), *Differential item functioning* (pp. 31-66). Erlbaum: Hillsdale NJ.
- Akcan, R. (2018). *2016 Lisans yerleştirme sınavı İngilizce testinin madde yanlılığı açısından incelenmesi* (Yüksek lisans tezi). Ankara: Hacettepe Üniversitesi.
- Akın Arkan, Ç., Uğurlu, S., & Atar, B. (2016). A DIF and bias study by using MIMIC, SIBTEST, Logistic Regression and Mantel-Haenszel methods. *Hacettepe University Journal of Education*, 31(1), 34-52.
- Arikan, S., van de Vijver, F., & Yagmur, K. (2018). Propensity score matching helps to understand sources of DIF and mathematics performance differences of Indonesian, Turkish, Australian, and Dutch students in PISA. *International Journal of Research in Education and Science (IJRES)*, 4(1), 69-81. doi:10.21890/ijres.382936
- Atalay, K. (2010). *PISA 2006 öğrenci anketinde yer alan tutum maddelerinin değişen madde fonksiyonu açısından incelenmesi* (Yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.
- Austin, P. C. (2013). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine*, 33(6), 1057–1069.
- Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: Guilford Press.
- Ayan, C. (2011). *PISA 2009 fen okuryazarlığı alt testinin değişen madde fonksiyonu açısından incelenmesi* (Yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.
- Bakan Kalaycıoğlu, D. & Kelecioğlu, H. (2011). Öğrenci Seçme Sınavı'nın madde yanlılığı açısından incelenmesi. *Eğitim ve Bilim*, 36(161), 3-13.

- Baser, O. (2006). Too much ado about propensity score models? Comparing methods of propensity score matching. *Value in Health*, 9(6), 377-385.
- Baykul, Y. (2015). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: Pegem Akademi.
- Brooks, G. P. & Johanson, G. A. (2003). Test analysis program. *Applied Psychological Measurement*, 27(4), 305-306. doi:[10.1177/0146621603027004007](https://doi.org/10.1177/0146621603027004007)
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. & Demirel, F. (2018). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi.
- Camilli, G. 2006. "Test Fairness". *Educational Measurement*. Editör: Brennan, R. L. Westport, CT: American Council on Education and Praeger.
- Camilli, G. & Shepard, L. A. (1994). *Methods for identifying biased test items*. London: Sage publications.
- Chen, M.Y., Liu, Y. & Zumbo, B.D. (2019). A propensity score method for investigating differential item functioning in performance assessment. *Educational and Psychological Measurement*, 80(3), 1-23. doi:10.1177/0013164419878861
- Clauser, B. E. & Mazor, K. M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice*, 17, 31-44.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, Inc.
- Çokluk, Ö. , Şekercioğlu, G. & Büyüköztürk, Ş. (2014). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL Uygulamaları*. Ankara: Pegem Akademi.
- D'Agostino, R. (1998). Tutorial in biostatistics propensity score methods for bias reduction in the comparison of a non- randomized control group, *Statistic in Medicine*, 17(19), 2265-2281.
- Demir, S. (2013). *Pisa 2009 matematik okuryazarlığı alt testinde bulunan maddelerinin Mantel-Haenszel, SIBTEST ve lojistik regresyon yöntemleri ile*

değişen madde fonksiyonunun incelenmesi (Yüksek lisans tezi). Abant İzzet Baysal Üniversitesi, Bolu.

Erdem, B. (2015). *Ortaöğretime geçişte kullanılan ortak sınavların değişen madde fonksiyonu açısından kitapçık türlerine göre farklı yöntemlerle incelenmesi* (Yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.

Erkuş, A. (2014). *Psikolojide ölçme ve ölçek geliştirme – I: Temel kavramlar ve işlemler*. Ankara: Pegem Akademi.

Fan, X., & Nowell, D. L. (2011). Using propensity score matching in educational research. *Gifted Child Quarterly*, 55(1), 74-79.

Ferrando, P.J. & Lorenzo-Seva, U. (2013). *Unrestricted item factor analysis and some relations with item response theory*. Technical Report. Department of Psychology, Universitat Rovira i Virgili, Tarragona.

Field, A. (2000). *Discovering Statistic Using SPSS for Windows*. Sage Publications Ltd.

Gök, B., Kelecioğlu, H., & Doğan, N. (2010). Değişen madde fonksiyonunu belirlemede Mantel Haenszel ve Lojistik Regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim Dergisi*, 35(156), 3-16.

Greifer, N. (2020). *cobalt: Covariate Balance Tables and Plots*. <https://cran.r-project.org/web/packages/cobalt/index.html> adresinden erişilmiştir.

Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: structures, distances, and algorithms. *Journal of Computational and Graphical Statistics*, 2(4), 405–420.

Gümüş, F. (2018). *TEOG matematik alt testinde değişen madde fonksiyonunun coğrafi bölgelere göre incelenmesi* (Yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.

Ho, D., Imai, K., King, G. & Stuart, E. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3), 199-236.

- Holland, P. W. & Wainer, H. (1993). *Differential item functioning*. United States of American Lawrence Erlbaum Associates, Inc.
- Holland, P. W. & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer ve H.I. Braun (Ed.), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.
- Jodoin, M. G. & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure ith the logistic regression procedure for DIF detection. *Applied Measurement in Education, 14*(4), 329-349.
- Joldersma, K., & Bowen, D. (2010, April). *Application of propensity models in DIF studies to compensate for unequal ability distributions*. Paper presented at the annual meeting of National Council on Measurement in Education, Denver, CO.
- Lee, H. & Geisinger, K. F. (2014) The effect of propensity scores on DIF analysis: inference on the potential cause of DIF, *International Journal of Testing, 14*(4), 313-338.
- Linn, R. L. (1993). *Educational assessment: Expanded expectations and challenge*. CSE Technical Report 351. CA: CRESST.
- Liu, Y., Zumbo, B. D., Gustafson, P., Huang, Y., Kroc, E., & Wu, A. D. (2016). Investigating causal DIF via propensity score methods. *Practical Assessment, Research & Evaluation, 21*(13), 1-24. doi: <https://doi.org/10.7275/ewqz-n963>
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers.
- Luellen, J. K., Shadish, W. R., & Clark, M. H. (2005). Propensity scores: An introduction and experimental test. *Evaluation Review, 29*(6), 530-558.
- Magis, D., Beland, S. & Raiche , G. (2016). *difR, version 4.7:Collection of methods to detect dichotomous differential item functioning(DIF)*. <https://cran.r-project.org/web/packages/difR/index.html> adresinden erişilmiştir.
- Martinková, P., Hladká, A., & Potužníková, E. (2020). Is academic tracking related to gains in learning competence? Using propensity score matching and

differential item change functioning analysis for better understanding of tracking implications. *Learning and Instruction*, 66, 101286.

MEB (2017). Akademik becerilerin izlenmesi ve değerlendirilmesi 8. sınıflar raporu. https://odsgm.meb.gov.tr/meb_iys_dosyalar/2017_11/30114819_iY-web-v6.pdf adresinden erişilmiştir.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-104). New York, NY: American Council on education and Macmillan.

Miller, M. D., Linn, R.L. & ve Gronlund, N. E. (2009). *Measurement and assessment in teaching* (10th ed.). Upper Saddle River, N.J. : Merrill/Pearson.

Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, 17(4), 297-334.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicers's MAP test. *Behavior Research Methods, Instruments & Computers*, 32(3), 396-402.

OECD (2019). *PISA 2018 Results (Volume I): What Student Know and Can Do, PISA*. Paris: OECD Publishing.

Olmos, A. & Govindasamy, P. (2015). Propensity scores: A practical introduction using R. *Journal of MultiDisciplinary Evaluation*. 11(25), 1-88.

Osterlind, J. S. (1983). *Test item bias*. London: Sage Publications.

Osterlind, J. S. (2002). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats* (2nd ed.). New York, Boston, Dordrecht, London, Moscow: Kluwer Academic Publishers.

Özmen, D. T. (2013). *PISA 2009 okuma becerileri testi maddelerinin yanlılık açısından Türkiye, Amerika Birleşik Devletleri ve Birleşik Krallık uygulamalarında karşılaştırılması* (Doktora tezi). Ankara Üniversitesi, Ankara.

Rogers, H. J., & Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement*, 17(2), 105–116.

- Rosenbaum, P. & Rubin, D. (1983) The central role of propensity score matching in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
- Rosenbaum, P. R., & Rubin, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society*, 45(2), 212–218.
- Rosenbaum, P. R., & Rubin, D.B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* 39(1), 33-38.
- Rosenbaum, P. R. (1986). Dropping out of high school in the United States: An observational study. *Journal of Educational Statistics* 11(3), 207-244.
- Rosenbaum, P. R. (2002). *Observational studies*. In *Observational studies* (pp. 1-17). Springer, New York, NY.
- Rosenbaum, P. R. (1989). Optimal matching for observational studies. *Journal of the American Statistical Association*, 84(408), 1024–1032.
- Rubin, D.B. (1997) Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, 127(8), 757-763.
- Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2, 169-188.
- Shepard, L. A. (1982). Definitions of bias. In R. A. Berk (Ed.), *Handbook of methods for detecting test bias* (pp. 9-30). Baltimore MD: The Johns Hopkins University Press.
- Stone, C. A. & Tang, Y. (2013). Comparing propensity score methods in balancing covariates and recovering impact in small sample educational program evaluations, *Practical Assessment, Research and Evaluation*, 18, Article 13.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1-21.
- Suna, H. E. (2012). *TIMSS 2007 fen bilimleri testindeki maddelerin dil ve cinsiyet yanlılığı açısından incelenmesi* (Yüksek lisans tezi). Ankara Üniversitesi, Ankara.

- Swaminathan, H. & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361-370.
- Tabachnick, B. G. & Fidell, L. S. (2013). *Using multivariate statistics* (Sixth Edition). United States: Pearson Education.
- Tekin, H. (2010). *Eğitimde ölçme ve değerlendirme* (17. Baskı). Ankara: Yargı Yayınevi.
- Thomas, N. & Rubin, D.B. (1996). Matching using estimated propensity scores: Relating theory to practice. *Biometrics*. 52(1), 249-264.
- Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research*, 46(1), 90-118.
- Thorndike, R. M., & Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education*. Boston: Prentice Hall.
- Turgut, F. & Baykul, H. (2015). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi Yayıncılık.
- Ulutaş, S. (2012). *PISA 2006 fen okuryazarlığı testindeki maddelerin yanlılık bakımından araştırılması* (Doktora tezi). Ankara Üniversitesi, Ankara.
- Van de Vijver, F. J. R. & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment. *European Review of Applied Psychology*, 54, 119-135.
- Wu, A. D., & Ercikan, K. (2006). Using multiple-variable matching to identify cultural sources of differential item functioning. *International Journal of Testing*, 6(3), 287-300.
- Wu, A.D., Liu, Y., Stone, J.E., Zou, D., & Zumbo, B.D. (2017). Is difference in measurement outcome between groups differential responding, bias or disparity? a methodology for detecting bias and impact from an attributional stance. *Frontiers in Education*, 2(39), 1-12. doi: 10.3389/feduc.2017.00039

- Yalçın, S. (2017). Öğrencilerin duyuşsal özelliklerine göre oluşan gizil sınıfların cinsiyete göre farklılaşan madde fonksiyonu'na etkisi. *İlköğretim Online*, 16(4), 1917-1931.
- Yıldırım, A. (2017). *PISA 2009 okuma becerileri alanındaki maddelerin tek deęişkenli ve çok deęişkenli eşleştirme yöntemi ile deęişen madde fonksiyonlarının incelenmesi* (Doktora Tezi). Ankara Üniversitesi, Ankara.
- Yıldırım, H. H. (2006). *The differential item functioning (DIF) analysis of mathematics items in the international assessment programs* (Doktora tezi). Ortadoęu Teknik Üniversitesi, Ankara.
- Yılmaz Koęar, E. (2019). The use of propensity score matching analysis in educational research. *Amasya Education Journal*, 8(1), 24-61.
- Yoshida, K. (2020). *tableone, version 0.12.0: Create 'Table 1' to describe baseline characteristics with or without propensity score weights*. <https://cran.r-project.org/web/packages/tableone/index.html> adresinden erişilmiştir.
- Zhang, Z., Kim, H.J., Lonjon, G., Zhu, Y. (2019). Balance diagnostics after propensity score matching. *Ann Transl Med.* 7(1), 16. doi: 10.21037/atm.2018.12.10
- Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P.W. Holland ve H. Wainer (Ed.), *Differential item functioning* (pp.337-347). Hillsdale, NJ: Erlbaum.
- Zumbo, B. D. & Thomas, D. R. (1996, October). *A measure of DIF effect size using logistic regression procedures*. Paper presented at the National Board of Medical Examiners, Philadelphia.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type (ordinal) item scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.
- Zumbo, B. D. & Gelin, M. N. (2005). A matter of test bias in educational policy research: bringing the context into picture by investigating sociological

community moderated (or mediated) test and item bias. *Journal of Educational Research and Policy Studies*, 5, 1-23.

Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4(2), 223-233.

EK-A: MEB Veri Talebi Cevap Yazısı



T.C.
MİLLÎ EĞİTİM BAKANLIĞI
Ölçme, Değerlendirme ve Sınav Hizmetleri
Genel Müdürlüğü

Sayı : 57750415-480.99-E.2711842

06.02.2020

Konu : ABİDE 2016 veri talebi

HACETTEPE ÜNİVERSİTESİ REKTÖRLÜĞÜNE
(Eğitim Bilimleri Enstitüsü Müdürlüğüne)

İlgi : 15.05.2019 tarih ve 51944218-300.E.00000592498 sayılı yazınız.

Eğitim Bilimleri Ana Bilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı tezli yüksek lisans programı öğrenciniz Mehmet YILMAZ'ın Dr. Öğr. Üyesi Kübra Atalay KABASAKAL'ın danışmanlığında yürüttüğü "Eğilim Puanları Kullanılarak ABİDE Çalışmasındaki Maddelerin Değişen Madde Fonksiyonu Açısından İncelenmesi" başlıklı tez çalışması kapsamında talep ettiği ABİDE-2016 verileri hazırlanmış olup CD ve veri teslim formu ile birlikte Mehmet YILMAZ'a elden teslim edilmiştir.

Bilgilerinizi ve gereğini rica ederim.

Dr. Sadri ŞENSOY
Bakan a.
Genel Müdür

EK- CD (1 adet)

(Not: CD elden teslim edilecektir)

EK-B: Etik Komisyonu Onay Bildirimi



T.C.
HACETTEPE ÜNİVERSİTESİ
Rektörlük

Tarih: 13.02.2019 14:57
Sayı: 35853172-300-E.00000458227

E.00000458227

Sayı : 35853172-300
Konu : Mehmet YILMAZ Hk.

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi : 01.02.2019 tarihli ve 51944218-300/00000441565 sayılı yazımız.

Enstitünüz Eğitim Bilimleri Ana Bilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı tezli yüksek lisans programı öğrencilerinden **Mehmet YILMAZ**'ın **Dr. Öğr. Üyesi Kübra ATALAY KABASAKAL** danışmanlığında hazırladığı "**Eğilim Puanları Kullanılarak Abide Çalışmasındaki Maddelerin Değişen Madde Fonksiyonu Açısından İncelenmesi**" başlıklı tez çalışması Üniversitemiz Senatosu Etik Komisyonunun **05 Şubat 2019** tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi ve gereğini saygılarımla rica ederim.

e-imzalıdır
Prof. Dr. Rahime Meral NOHUTCU
Rektör Yardımcısı

Evrakın elektronik imzalı suretine <https://belgedogrulama.hacettepe.edu.tr> adresinden 37ae783f-8f92-4f66-a2ca-695d8360004e kodu ile erişebilirsiniz. Bu belge 5070 sayılı Elektronik İmza Kanunu'na uygun olarak Güvenli Elektronik İmza ile imzalanmıştır.

Hacettepe Üniversitesi Rektörlük 06100 Sıhhiye-Ankara
Telefon:0 (312) 305 3001-3002 Faks:0 (312) 311 9992 E-posta:yazimd@hacettepe.edu.tr İnternet
Adresi: www.hacettepe.edu.tr

Duygu Didem İLFRİ

